

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 26, 2004, 19:04:49 ; Search time 73 Seconds  
(without alignments)  
28.266 Million cell updates

Title: US-09-766-412-30

Perfect score: 66  
Sequence: 1 QPVLHLVALNTPL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB sec length: 0

Maximum DB seq length: 2000000000

Post-processing. Minimum. March 08

Post-processing:	Minimum	Match	0%
	Maximum	Match	100%

Listing first 45 summaries

Database : A Geneseq 19Jun03:\*

1:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA2002.DAT.*
24:	/SIDSI/cgdata/geneseq/geneseqp-emb1/AA2003.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query #	Query			ID	Description
		Score	Match	Length		
1	66	100.0	13	22	AAB80862	Angiostatin-derive
2	66	100.0	13	22	AAB74286	Exemplary anti-ang
3	66	100.0	13	23	ABG97533	Antiangiogenic pep
4	66	100.0	20	18	AAB16596	N-terminal region
5	66	100.0	20	21	AA194381	Mouse endostatin N
6	66	100.0	20	23	AAB48821	Endostatin N-termi
7	66	100.0	22	23	AAB49806	Murine endostatin
8	66	100.0	39	22	AAB35594	Antiangiogenic non
9	66	100.0	42	22	AAB71931	Murine sig-mendo f

10	56	100.0	183	23	AAW495004
11	66	100.0	184	20	AAV18409
12	66	100.0	184	20	AAV08689
13	66	100.0	184	21	AAV70258
14	66	100.0	184	21	AAW49380
15	66	100.0	184	23	ABG31793
16	66	100.0	184	23	AAV06197
17	66	100.0	191	21	AAE28398
18	66	100.0	191	23	AAV77950
19	66	100.0	207	22	AAE02031
20	66	100.0	207	22	AAE71930
21	66	100.0	207	23	ABE79302
22	66	100.0	218	20	AAV08691
23	66	100.0	580	20	AAE08692
24	66	100.0	684	20	AAV25114
25	66	100.0	1288	18	AAW26328
26	66	100.0	1288	20	AAW92397
27	62	93.9	20	22	AAE49850
28	62	93.9	25	22	ABE49856
29	62	93.9	44	23	ABG31787
30	62	93.9	44	23	ABG31791
31	62	93.9	178	21	AAV94324
32	62	93.9	178	22	AAU00900
33	62	93.9	179	22	AAU00901
34	62	93.9	180	22	AAU00899
35	62	93.9	180	22	AAU00898
36	62	93.9	182	21	AAE28398
37	62	93.9	182	21	AAV94323
38	62	93.9	182	21	AAV59622
39	62	93.9	182	22	AAU00897
40	62	93.9	182	23	AAU77951
41	62	93.9	183	20	AAV08693
42	62	93.9	183	20	AAI02113
43	62	93.9	183	21	AAE03493
44	62	93.9	183	21	AAE16451
45	62	93.9	183	21	AAV90771

## ALIGNMENTS

RESULT 1  
AAB80862  
ID AAB80862 standard; Peptide; 13 AA.

AAC AAB80862;

29-MAY-2001 (first entry)

Angiostatin-derived peptide #19.

Cytostatic; angiogenesis inhibitor; tumour metastasis; tumour growth.

Unidentified.

XX  
PN  
US6200954-B1.

13-MAR-2001.

XX  
PF 30-AUG-1999:

XX  
PR 04-SEP-1998: 98US-0099313.

AA (UYSI-) UNIV SINGAPORE NAT.

XX  
PT Ge R. Kini RM:XX  
DR WPI: 2001-234520

Peptides having potent

XX  
XX  
TREATMENT OF CANCORS, COMPRISE A PORTION OF A LITERATURE PROCEEDING

PS Disclosure; Columns 13-16; zipp; english.

XX The present sequence is a peptide which is effective in inhibiting  
 CC undesirable angiogenesis. Angiogenesis is the process of new blood vessel  
 CC formation from pre-existing vessels. Inappropriate angiogenesis is  
 CC associated with various pathological conditions including solid tumour  
 CC growth and metastasis. The present peptide can be used to prevent  
 CC tumour metastasis or inhibit the growth of a primary tumour.

XX SQ Sequence 13 AA;

Query Match 100.0%; Score 66; DB 22; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPVLHLVALNTPPL 13  
 |||||  
 Db 1 QPVLHLVALNTPPL 13

RESULT 2

AAAB74256  
 ID AAB74256 standard; peptide; 13 AA.

XX AC AAB74256;

DT 23-MAY-2001 (first entry)

XX Exemplary anti-angiogenic peptide #30.

XX Anti-angiogenesis; plasminogen; VEGF;  
 KW vascular endothelial growth factor; FLT-1; tumour; metastasis;  
 KW cancer.

XX Homo sapiens.

XX WO200118030-A2.

XX 15-MAR-2001.

XX 01-SEP-2000; 2000WO-SG00131.

XX 03-SEP-1999; 99SG-0004310.

XX (UYSI-) UNIV SINGAPORE NAT.

XX Ge R, Kini RM;

XX WPI; 2001-257785/26.

XX Peptides comprising a portion of a protein selected from plasminogen,  
 PT endostatin, VEGF, FLT-1 and KDR/FLK-1 are useful for treating primary  
 PT tumor growth.

XX Claim 7; Page 18; 34pp; English.

XX The present invention relates to anti-angiogenesis peptides from  
 CC a portion of a selected from plasminogen, endostatin, VEGF or FLT-1.  
 CC The invention is used to prevent or treat primary tumour growth or  
 CC metastasis or undesired angiogenesis.

XX SQ Sequence 13 AA;

Query Match 100.0%; Score 66; DB 22; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPVLHLVALNTPPL 13  
 |||||  
 Db 1 QPVLHLVALNTPPL 13

RESULT 3

ABG97539

ID ABG97539 standard; Peptide; 13 AA.

XX ABG97539;

DT 16-DEC-2002 (first entry)

XX Antiangiogenic peptide, Endo-2.

XX Angiogenesis; plasminogen; endostatin; Endo-; VEGF;

KW vascular endothelial growth factor; FLT-1; kinase insert domain; FLK-1;  
 KW angiogenic inhibitor; Angio-; endothelial cell; proliferation;

KW tumour growth; blood vessel formation; migration; tubule structure;

KW embryonic development; wound healing; tumour metastasis;

KW rheumatoid arthritis; psoriasis; anticancer; therapy;

KW antiangiogenic therapy; mitogen; tyrosine kinase receptor.

XX Mammalia.

XX US2002103129-A1.

XX 01-AUG-2002.

XX 22-JAN-2001; 2001US-0766412.

XX 04-SEP-1998; 98US-099313P.

XX 30-AUG-1999; 99US-0385442.

XX (GERR/) GE R.

XX (KINI/) KINI R M.

XX Ge R, Kini RM;

XX WPI; 2002-749595/81.

XX Claim 9; Page 10; 24pp; English.

XX The invention discloses a peptide comprising a portion of a protein e.g.  
 CC plasminogen, endostatin (Endo-1-4), vascular endothelial growth factor  
 CC (VEGF), FLT-1 or kinase insert domain containing receptor/FLK-1. These  
 CC peptides are able to act as potent angiogenic inhibitors (Angio-1-5)  
 CC inhibiting endothelial cell proliferation and retarding tumour growth.

XX Angiogenesis is the process of new blood vessel formation from  
 CC pre-existing vessel, involving endothelial cell proliferation, migration  
 CC and assembly into tubule structures. It plays an important role in normal  
 CC physiological functions such as embryonic development and wound healing.

XX Inappropriate angiogenesis is also associated with various pathological  
 CC conditions including tumour growth and metastasis, rheumatoid arthritis  
 CC and psoriasis. Anticancer therapy by inhibiting tumour angiogenesis is  
 CC called antiangiogenic therapy. VEGF is an endothelial specific mitogen  
 CC that functions through two high affinity tyrosine kinase receptors, FLT-1  
 CC and FLK-1 (not defined). Protein-protein interactions are crucial to many  
 CC physiological and pharmacological processes and the peptides disclosed  
 CC interfere with these interactions. The peptides are useful for preventing  
 CC or treating undesired angiogenesis and primary tumour growth or  
 CC metastasis. The sequences presented in ABG97510-ABG97559 are examples of  
 CC antiangiogenic peptides.

XX SQ Sequence 13 AA;

Query Match 100.0%; Score 66; DB 23; Length 13;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPVLHLVALNTPPL 13  
 |||||  
 Db 1 QPVLHLVALNTPPL 13

## RESULT 4

AAW16596  
ID AAW16596 standard; peptide; 20 AA.  
XX  
AC AAW16596;  
XX  
AC AAW16596;  
XX  
DT 28-JAN-1998 (first entry)  
XX  
DE N-terminal region of a novel endostatin.  
XX  
DE Endostatin; angiogenesis; cancer; tumour; rheumatoid arthritis;  
XX psoriasis; ocular; Osler-Webber Syndrome; myocardial; telangiectasia;  
XX plaque neovascularisation; haemophilic joint; angiofibroma; inhibitor;  
XX wound granulation; intestinal adhesion; atherosclerosis; scleroderma;  
XX hypertrophic scar; cat scratch disease; Rochelle minialia quintosa;  
XX Helicobacter pylori ulcer; birth control; collagen alpha type XVIII.  
XX  
OS Mus sp.  
XX  
XX WO9715666-A1.  
XX  
XX PD 01-MAY-1997.  
XX  
XX PF 23-OCT-1996; 96WO-US16925.  
XX  
XX PR 22-OCT-1996; 96US-0740168.  
XX PR 23-OCT-1996; 96US-0005835.  
XX PR 02-AUG-1996; 96US-0023075.  
XX PR 17-SEP-1996; 96US-0026263.  
XX  
XX PA (CHIL-) CHILDRENS MEDICAL CENT.  
XX  
XX PI Folkman MJ, O'Reilly MS;  
XX  
XX WPI; 1997-259020/23.  
XX  
XX PT Isolated endostatin - useful for treating, e.g. angiogenesis  
XX dependent cancers  
XX  
XX PS Claim 3; Page 56; 84pp; English.  
XX  
XX CC A novel 20kDa endostatin that specifically inhibits endothelial  
XX cell proliferation was isolated and is characterised by its preferred  
XX N-terminal amino acid sequence. The present sequence represents the  
XX first 20 of these amino acids. The N-terminal sequence corresponds  
XX to an internal 20 amino acid fragment found in mouse collagen alpha type  
XX XVIII satting at amino acid 1105 and ending at amino acid 1124. The  
XX N-terminal amino acid sequence of the inhibitor also corresponds to an  
XX internal 20 amino acid fragment found in human alpha 1 type XVIII  
XX starting at amino acid 1132 and ending at amino acid 1151. Endostatin  
XX can be isolated from murine hemangioendothelioma. The endostatin can be  
XX used to treat angiogenesis related diseases, e.g. angiogenesis dependent  
XX cancers, benign tumours, rheumatoid arthritis, psoriasis, ocular  
XX angiogenesis diseases, Osler-Webber Syndrome, myocardial angiogenesis,  
XX plaque neovascularisation, telangiectasia, haemophilic joints,  
XX angiofibroma, wound granulation, intestinal adhesions, atherosclerosis,  
XX scleroderma, hypertrophic scars, cat scratch disease (Rochelle minialia  
XX quintosa) and Helicobacter pylori ulcers. The endostatin can also be used  
XX to prevent embryo implantation, i.e. in birth control.  
XX  
XX SQ Sequence 20 AA;

Query Match 100.0%; Score 66; DB 18; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.4e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
| | | | | | | | | | | | | |  
Db 7 QPVLHLVALNTPL 19

## RESULT 5

AAW94321

AAW94321 standard; peptide; 20 AA.  
XX  
AC AAW94321;  
XX  
DT 11-AUG-2000 (first entry)  
XX  
DE Mouse endostatin N-terminus.  
XX  
XX Mouse; endothelial cell proliferation inhibitor; collagen XVIII;  
XX angiogenesis inhibitor; anti-tumour; cytostatic; antipsoriatic;  
XX vasotropic; dermatological; ophthalmological; vulnary;  
XX antiarteriosclerotic; antidiabetic; haemostatic; contraceptive;  
XX ocular angiogenic disease; atherosclerosis; scleroderma;  
XX myocardial angiogenesis; telangiectasia; angiofibroma;  
XX wound granulation.  
XX  
OS Mus musculus.  
XX  
XX WO200026368-A2.  
XX  
XX PD 11-MAY-2000.  
XX  
XX PF 01-NOV-1999; 99WO-US25605.  
XX  
XX PR 30-OCT-1998; 98US-0106343.  
XX PR 20-MAY-1999; 99US-0315689.  
XX  
XX PA (CHIL-) CHILDRENS MEDICAL CENT.  
XX  
XX PI O'Reilly MS, Folkman MJ;  
XX  
XX WPI; 2000-365617/31.  
XX  
XX PT Novel endostatin capable of inhibiting endothelial cell proliferation  
XX and angiogenesis, useful for treating angiogenesis-dependent cancers  
XX and as birth control agents  
XX  
XX PS Disclosure; Page 2; 68pp; English.  
XX  
XX CC The present sequence is the N-terminus of an endostatin, a potent and  
XX specific protein inhibitor of endothelial cell proliferation in  
XX vitro and angiogenesis in vivo. The sequence corresponds to a  
XX C-terminal fragment of newly identified collagen type XVIII.  
XX CC Recombinant mouse endostatin (20 mg/kg) was administered subcutaneously  
XX to mice implanted with Lewis lung carcinomas. There was tumour mass  
XX regression non-detectable levels after 12 days of therapy due to the  
XX angiogenesis inhibitory activity of endostatin. Thus the protein is  
XX useful for treatment of angiogenesis-dependent cancers. The  
XX polynucleotide and polypeptide sequences of this endostatin are useful  
XX for treating and diagnosis of tumours, ocular angiogenic diseases,  
XX Osler-Webber syndrome, myocardial angiogenesis, plaque  
XX neovascularisation, telangiectasia, haemophilic joints, angiofibroma  
XX and wound granulation, for treatment of diseases related to excessive or  
XX abnormal stimulation of endothelial cells e.g. intestinal adhesions,  
XX atherosclerosis, scleroderma. The protein may also be useful as a birth  
XX control agent by reducing or preventing uterine vascularisation. The  
XX gene for endostatin may be isolated from cells or tissue that express  
XX high levels of endostatin, eg. tumour cells, by generating cDNA from  
XX mRNA using reverse transcriptase and then amplifying the DNA sequence.  
XX  
XX SQ Sequence 20 AA;

Query Match 100.0%; Score 66; DB 21; Length 20;  
Best Local Similarity 100.0%; Pred. No. 3.4e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
| | | | | | | | | | | | | |  
Db 7 QPVLHLVALNTPL 19

RESULT 6  
AAW48821

ID AAM48821 standard; Peptide; 20 AA.  
 XX  
 AC AAM48821;  
 XX  
 DT 04-APR-2002 (first entry)  
 XX  
 DE Endostatin N-terminal peptide.  
 XX  
 KW Human; angiotensin; endostatin; angiogenesis; cancer; metastasis;  
 KW psoriasis; scleroderma; Crohn's disease; corneal disease;  
 KW retinopathy; arthritis; wound healing; Helicobacter pylori; peptic ulcer;  
 KW gene therapy; angiotensin antagonist; endostatin antagonist;  
 KW antiangiogenic; cytostatic; antiarthritic; antiinflammatory;  
 KW cerebroprotective; antidiabetic; virucide; antipyretic; vulnery;  
 KW gynaecological; cat scratch fever.  
 XX  
 OS Unidentified.  
 XX  
 PN WO200193897-A2.  
 XX  
 PD 13-DEC-2001.  
 XX  
 PF 04-JUN-2001; 2001WO-US17947.  
 XX  
 PR 02-JUN-2000; 2000US-209065P.  
 PR 08-MAY-2001; 2001US-289387P.  
 XX  
 PA (ENTR-) ENTREMED INC.  
 XX  
 PI Sim XL, MacDonald NJ;  
 XX  
 DR WPI; 2002-130569/17.  
 XX  
 PR Regulating angiogenesis and treatment of angiogenesis-mediated  
 PT diseases, e.g. hemangioma, tumors or cancer, by administering a  
 PT tropomyosin binding compound or actin disrupting compound -  
 XX  
 PS Disclosure; Page 13; 95pp; English.  
 XX  
 CC The present invention relates to methods of regulating angiogenesis in an  
 CC individual by administering an angiogenesis regulating composition  
 CC comprising a tropomyosin binding compound or an actin disrupting  
 CC compound. The compositions are useful for treating diseases and processes  
 CC mediated by angiogenesis including haemangioma, solid tumours, blood  
 CC borne tumours, leukaemia, metastasis, Crohn's disease, coronary or  
 CC cerebral collateral, arthritis, diabetic neovascularisation, macular  
 CC degeneration, wound healing, Helicobacter related diseases, ovulation,  
 CC menstruation, and cat scratch fever. The present sequence is a peptide  
 CC described in the exemplification of the invention.  
 XX  
 SQ Sequence 20 AA;  
 Query Match 100.0%; Score 66; DB 23; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 3.4e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QPVLHLVALNTPL 13  
 DB 7 QPVLHLVALNTPL 19  
 RESULT 7  
 AAB49806  
 ID AAB49806 standard; Peptide; 23 AA.  
 XX  
 AC AAB49806;  
 XX  
 DT 02-MAR-2001 (first entry)  
 XX  
 DE Murine endostatin peptide fragment SEQ ID NO: 19.  
 XX  
 KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;  
 KW cancer; inflammation; angiogenesis-dependent disease.

XX Mus musculus.  
 OS  
 PN WO200067771-A1.  
 XX  
 PD 16-NOV-2000.  
 XX  
 PF 02-MAY-2000; 2000WO-US12063.  
 XX  
 PR 06-MAY-1999; 99US-0132907.  
 PR 14-JUL-1999; 99US-0353333.  
 XX  
 PA (BURN-) BURNHAM INST.  
 XX  
 PI Vuori K;  
 XX  
 DR WPI; 2001-040937/05.  
 XX  
 PT Endostatin peptide comprising at least four endostatin amino acid  
 PT residues are e.g. angiogenesis inhibitors for treating cancer and  
 PT diabetic retinopathy -  
 XX  
 PS Example 4; Page 124; 145pp; English.  
 XX  
 CC The present invention provides endostatin peptides which can be used in  
 CC the modulation of angiogenesis. This is useful in the treatment of  
 CC cancers, inflammation, rheumatoid arthritis, chronic articular  
 CC rheumatism, psoriasis, disorders associated with inopportune invasion of  
 CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy  
 CC of prematurity, macular degeneration, corneal graft rejection,  
 CC retrolental fibroplasia, rubeosis, capillary proliferation in,  
 CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent  
 CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque  
 CC neovascularisation, telangiectasia, haemophilic joints and wound  
 CC granulation. In addition, the peptides can be used as birth control  
 CC agents.  
 XX  
 SQ Sequence 23 AA;  
 Query Match 100.0%; Score 66; DB 22; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-05;  
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QPVLHLVALNTPL 13  
 DB 7 QPVLHLVALNTPL 19  
 RESULT 8  
 AAB35584  
 ID AAB35584 standard; peptide; 39 AA.  
 XX  
 AC AAB35584;  
 XX  
 DT 14-FEB-2001 (first entry)  
 XX  
 DE Antiangiogenic nonatriacapeptide I.  
 XX  
 KW Antiangiogenic; angiogenesis; cancer; endostatin.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200063249-A1.  
 XX  
 PD 26-OCT-2000.  
 XX  
 PF 11-APR-2000; 2000WO-EF03236.  
 XX  
 PR 15-APR-1999; 99IT-MI00777.  
 XX  
 PA (UYMI-) UNIV MILANO.  
 PA (UYFI-) UNIV FIRENZE.  
 XX

PI Chillemi F, Francescato P, Ziche M;  
DR WPI; 2001-007005/01.  
XX  
XX Polypeptides derived from endostatin exhibiting antiangiogenic activity  
PT useful for treatment of angiogenesis-dependent tumours  
XX  
XX Claim 2; Page 18; 28pp; English.  
PS  
XX The present invention describes a number of peptides derived from  
CC endostatin which exhibit antiangiogenic activity. These may be used in  
CC the treatment of cancer. The present sequence is one of the peptides of  
CC the invention.  
XX  
XX SQ Sequence 39 AA;  
Query Match 100.0%; Score 66; DB 22; Length 39;  
Best Local Similarity 100.0%; Pred. No. 7.2e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QPVLHLVALNTPL 13  
Db 7 QPVLHLVALNTPL 19  
RESULT 9  
AAB71931  
ID AAB71931 standard; Protein; 42 AA.  
XX  
XX AAB71931;  
AC  
XX  
XX 10-MAY-2001 (first entry)  
DT  
XX  
XX Murine sig-mEndo fusion protein N-terminal sequence.  
DE  
XX  
XX Mouse; endostatin; antitumor; cytostatic; antiarthritic; antipsoriatic;  
KW antidiabetic; ophthalmological; gene therapy; angiogenic inhibitor;  
KW adenoviral vector; diabetic retinopathy; cardiovascular disease;  
KW arthritis; psoriasis; cerebral oedema; intravascular coagulopathy;  
KW lymphoma; leukaemia; sig-mEndo; fusion protein.  
XX  
XX OS Mus sp.  
XX  
XX WO200112830-A1.  
FN  
XX  
XX 22-FEB-2001.  
PD  
XX  
XX 11-AUG-2000; 2000WO-EP07865.  
PF  
XX  
XX 13-AUG-1999; 99US-0373938.  
PR  
XX  
XX (NOVS) NOVARTIS AG.  
PA  
XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.  
XX  
PI Hallenbeck PL, Chen CT;  
XX  
XX WPI; 2001-202871/20.  
DR  
XX  
XX Adenoviral vector for treating tumors and disorders associated with  
PT angiogenesis, such as cancer, arthritis, and psoriasis, comprises a DNA  
PT sequence encoding an angiogenic inhibitor, particularly endostatin -  
XX  
XX Example 1; Fig 5; 59pp; English.  
PS  
XX The present peptide was analysed in a study of the expression and  
CC analysis of sig-mEndo, which comprises murine endostatin attached to the  
CC murine Ig-kappa signal sequence. sig-mEndo was introduced into an  
CC adenoviral vector. The adenoviral vector is useful for expressing  
CC endostatin in a mammalian cell such as an A549 or Hep3B cell. It is  
CC useful for treating other diseases and disorders associated with  
CC angiogenesis, such as neovascular diseases of the eye, including diabetic  
CC retinopathy, cardiovascular disease, arthritis, psoriasis, cerebral  
CC oedema and intravascular coagulopathy (Kasabach-Merritt syndrome). The

CC vector inhibits, prevents or destroys the growth of tumours by  
CC preventing the formation of blood vessels in tumours, such as lymphoma  
CC and leukaemia.  
XX  
XX SQ Sequence 42 AA;  
Query Match 100.0%; Score 66; DB 22; Length 42;  
Best Local Similarity 100.0%; Pred. No. 7.8e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QPVLHLVALNTPL 13  
Db 30 QPVLHLVALNTPL 42

RESULT 10  
AAM49504  
ID AAM49504 standard; Protein; 183 AA.  
XX  
XX AAM49504;  
AC  
XX  
XX 07-MAY-2002 (first entry)  
DT  
XX  
XX Mouse endostatin protein.  
DE  
XX  
XX Endostatin; murine; proliferation; blood vessel endothelium;  
KW regeneration; tumour; blood vessel; treatment; amplification.  
KW  
XX  
XX OS Mus sp.

XX  
XX Key Location/Qualifiers  
FH Misc-difference 181  
FT FT /label= Ser, Phe  
FT  
XX  
XX CN1177005-A.  
FN  
XX  
XX 25-MAR-1998.  
PD  
XX  
XX 10-SEP-1997; 97CN-0107112.  
PF  
XX  
XX 10-SEP-1997; 97CN-0107112.  
PR  
XX  
XX (XUGG/) XU G.  
PA  
XX  
XX Xu G, Ren M, Xu L;  
PI  
XX  
XX WPI; 2002-106746/15.  
DR  
XX  
XX Gene clone of inhibitory factor for hyperplasia of inner blood vessel  
PT cells in human body's real tumor, and its use in anti-tumor blood  
PT vessel regeneration -  
XX  
XX Disclosure; Page 4 (Disclosure); 6pp; Chinese.

XX  
XX This invention describes a novel preparation which inhibits the  
CC proliferation of blood vessel endothelium and prevents the regeneration  
CC activity of tumour blood vessels. The preparation can also be used as a  
CC biological preparation in the treatment of tumours. This sequence  
CC represents the murine endostatin protein described in the invention.  
XX  
XX SQ Sequence 183 AA;  
Query Match 100.0%; Score 66; DB 23; Length 183;  
Best Local Similarity 100.0%; Pred. No. 0.00041;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QPVLHLVALNTPL 13  
Db 7 QPVLHLVALNTPL 19  
RESULT 11  
AAV18409

ID AAY18409 standard; Protein; 184 AA.  
XX  
AC AAY18409;  
XX  
DT 24-AUG-1999 (first entry)  
XX  
DE Endostatin protein sequence.  
XX  
KW EMI; anti-angiogenic peptide; endostatin; angiogenesis-dependent cancer;  
KW benign tumour; rheumatoid arthritis; psoriasis; ocular angiogenesis;  
KW Osler-Webber Syndrome; myocardial angiogenesis; angiofibroma; cancer;  
KW plaque neovascularisation; telangiectasia; atherosclerosis; scleroderma;  
KW dialysis graft vascular access stenosis; renal cancer; therapy.  
XX  
OS Mus sp.  
XX  
XX WO9929855-A1.  
XX  
XX 17-JUN-1999.  
XX  
XX 08-DEC-1998; 98WO-US26057.  
XX  
XX 16-NOV-1998; 98US-0108536.  
PR 08-DEC-1997; 97US-0067888.  
PR 22-APR-1998; 98US-0082663.  
XX  
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
XX  
XX Sukhatme VP;  
PI  
XX WPI; 1999-385604/32.  
DR N-PSDB; AAX79949.  
XX  
XX Mutant endostatin having anti-angiogenic activity  
PT  
PS Claim 31; Fig 2; 105pp; English.  
XX  
XX This sequence is the mouse endostatin. The invention relates to a  
CC the mutant endostatin (EMI), which has anti-angiogenic activity, and is  
CC designated EMI. Compositions comprising EMI or fusion proteins comprising  
CC EMI, are useful for treating diseases characterised by angiogenic  
CC activity, such as angiogenesis-dependent cancers, benign tumours,  
CC rheumatoid arthritis, psoriasis, ocular angiogenesis, Osler-Webber  
CC Syndrome, myocardial angiogenesis, plaque neovascularisation,  
CC telangiectasia, haemophilic joints, angiofibroma, wound granulation,  
CC intestinal adhesions, atherosclerosis, scleroderma, hypertrophic scars,  
CC cat scratch disease, Helicobacter pylori ulcers, dialysis graft vascular  
CC access stenosis, contraception and obesity. In particular, the diseases  
CC treatable by EMI comprise cancer, especially renal cancer. The methods  
CC provide a means for introducing EMI into mammalian cells via gene  
CC therapy, for production of EMI via recombinant means, as well as  
CC recombinant production of the EMI protein. EMI performs as well or better  
CC than whole endostatin. Use of EMI is advantageous for treatment of  
CC angiogenic diseases in that increasingly smaller peptides are more potent  
CC on a weight basis, and may be able to better penetrate tissues.  
XX  
XX Sequence 184 AA;  
SQ  
Query Match 100.0%; Score 66; DB 20; Length 184;  
Best Local Similarity 100.0%; Pred. No. 0.00041;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QPVLHLVALNTPL 13  
DB 7 QPVLHLVALNTPL 19  
RESULT 12  
AAY08689  
ID AAY08689 standard; Protein; 184 AA.  
XX  
XX AAY08689;  
AC  
XX

DT 10-AUG-1999 (first entry)  
XX  
DE Murine endostatin protein fragment.  
XX  
KW Plasminogen; murine; angiotatin; endostatin; gene therapy; vector;  
KW anti-angiogenic; attenuation; cytostatic; anti-diabetic; ophthalmology;  
KW tumour growth; solid tumour; diabetic retinopathy; retina.  
XX  
OS Mus sp.  
XX  
XX WO9926480-A1.  
PN  
XX 03-JUN-1999.  
PD  
XX 20-NOV-1998; 98WO-US24950.  
PF  
XX 20-NOV-1997; 97US-0975424.  
PR  
XX (GENE-) GENETIX PHARM INC.  
PA (MASI) MASSACHUSETTS INST TECHNOLOGY.  
XX  
XX Bachelot T, Leboulch P, Pawliuk RJ;  
PI  
XX WPI; 1999-357696/30.  
DR N-PSDB; AAX77715.  
DR  
XX Anti-angiogenic gene therapy vectors  
PT  
XX Disclosure; Fig 6; 83pp; English.  
PS  
XX This invention describes a novel viral gene therapy vector comprising a  
CC nucleic acid molecule encoding an anti-angiogenic polypeptide chosen  
CC from human or murine angiotatin, human or murine endostatin and  
CC angiogenesis-inhibiting fusions and fragments, where the viral vector is  
CC sufficiently attenuated for use in human gene therapy. The products of  
CC the invention have anti-angiogenic, cytostatic, anti-diabetic and  
CC ophthalmological activity. The vector is used in gene therapy for  
CC inhibiting tumour growth in humans harbouring a solid tumour. The vector  
CC expresses an anti-angiogenic polypeptide. An additional use comprises  
CC treatment of diabetic retinopathy, where the anti-angiogenic polypeptide  
CC inhibits angiogenesis in the vicinity of the retina. The vector is  
CC administered to cells ex vivo and then administered to the patient.  
XX  
XX Sequence 184 AA;  
SQ  
Query Match 100.0%; Score 66; DB 20; Length 184;  
Best Local Similarity 100.0%; Pred. No. 0.00041;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 QPVLHLVALNTPL 13  
DB 7 QPVLHLVALNTPL 19  
RESULT 13  
AAY70258  
ID AAY70258 standard; Protein; 184 AA.  
XX  
XX AAY70258;  
AC  
XX 06-JUN-2000 (first entry)  
DT  
XX Murine angiogenesis inhibitor, endostatin.  
DE  
XX Murine; immunoglobulin Fc fragment; endostatin; immunofusin;  
KW angiogenesis; inhibitor; cytostatic; antirheumatoid; antiarthritic;  
KW antipsoriatic; antidiabetic; ophthalmological; immunosuppressant;  
KW vasotropic; vulvular; treatment; antiarteriosclerosis; tumour;  
KW metastasis; atherosclerosis; psoriasis; rheumatoid arthritis;  
KW ocular angiogenic disease; diabetic retinopathy; macular degeneration;  
KW myocardial angiogenesis; plaque neovascularisation; telangiectasia;  
KW wound granulation; keloid scar; gene therapy.  
XX

OS Mus musculus.

XX  
XX WO200011033-A2.  
XX  
XX PD 02-MAR-2000.  
XX  
XX PF 25-AUG-1999; 99WO-US19329.  
XX  
XX PR 25-AUG-1998; 98US-0097883.  
XX  
XX PA (LEXI-) LEXINGEN PHARM CORP.  
XX  
XX PI Lo K, Li Y, Gillies SD;  
XX WPI; 2000-237616/20.  
XX DR N-PSDB; AAZS1299.  
XX

PT Novel fusion protein of angiostatin or endostatin and an immunoglobulin  
PT FC region, useful for treating conditions mediated by angiogenesis,  
PT such as rheumatoid arthritis, tumors and macular degeneration -  
XX  
XX Example 5; Pages 48-49; 68pp; English.

XX The patent discloses a DNA molecule encoding a fusion protein comprising  
CC a signal sequence, an immunoglobulin Fc region, and an angiogenesis  
CC inhibitor selected from angiostatin, endostatin, a plasminogen fragment  
CC having angiostatin activity, a collagen XVIII fragment having endostatin  
CC activity, or combinations of them. The fusion protein (immunofusin) is  
CC used to inhibit angiogenesis and to treat diseases or conditions mediated  
CC by angiogenesis. Conditions that may be treated include solid tumours,  
CC blood born tumours, tumour metastasis, benign tumours including  
CC haemangiomas, acoustic neuromas, neurofibromas, trachomas and pyrogenic  
CC granulomas, rheumatoid arthritis, psoriasis, ocular angiogenic diseases  
CC e.g. diabetic retinopathy, retinopathy of prematurity, macular  
CC degeneration, corneal graft rejection, neovascular glaucoma, retrolental  
CC fibroplasia, rubeosis and Osler-Webber syndrome; myocardial angiogenesis,  
CC plaque neovascularisation, telangiectasia, haemophilic joints,  
CC angiofibroma, wound granulation, and excessive or abnormal stimulation of  
CC endothelial cells, intestinal cells, atherosclerosis, sclerodermal and  
CC hypertrophic scars i.e. keloid scars. The DNA constructs may be used  
CC in gene therapy. The present sequence is a murine  
CC endostatin used in the construction of immunofusin containing murine  
CC immunoglobulin Fc fragment.  
XX  
XX SQ Sequence 184 AA;

Query Match 100.0%; Score 66; DB 21; Length 184;  
Best Local Similarity 100.0%; Pred. No. 0.00041;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVHLHVALNTPL 13  
Db 7 QPVHLHVALNTPL 19  
|||||  
|||||

RESULT 14  
AAB49380  
ID ID AAB49380 standard; Protein; 184 AA.  
XX  
XX AC AAB49380;  
XX  
XX DT 02-MAR-2001 (first entry)  
XX  
XX DE Murine endostatin SEQ ID NO: 4.  
XX  
XX KW Endostatin; antiangiogenic; angiogenesis; human; mouse; chicken;  
KW cancer; inflammation; angiogenesis-dependent disease.  
XX  
XX OS Mus musculus.  
XX  
XX PN WO200067771-A1.  
XX  
XX PD 16-NOV-2000.

```

XX 02-MAY-2000; 2000WO-US12063.
XX
XX 06-MAY-1999; 99US-0132907.
XX PR 14-JUL-1999; 99US-0353333.
XX
XX (BURN-) BURNHAM INST.
XX
XX PA Vuori K;
XX PI
XX
XX NPI; 2001-040937/05.
XX DR N-PSDB; AAC88230.
XX
XX
XX Endostatin peptide comprising at least four endostatin amino acid
PT residues are e.g. angiogenesis inhibitors for treating cancer and
PT diabetic retinopathy -
XX
XX Disclosure; Fig 1; 146pp; English.
XX
XX The present invention provides endostatin peptides which can be used in
CC the modulation of angiogenesis. This is useful in the treatment of
CC cancers, inflammation, rheumatoid arthritis, chronic articular
CC rheumatism, psoriasis, disorders associated with inopportune invasion of
CC vessels such as diabetic retinopathy, neovascular glaucoma, retinopathy
CC of prematurity, macular degeneration, corneal graft rejection,
CC retrolental fibroplasia, rubeosis, capillary proliferation in
CC atherosclerotic plaques and osteoporosis. Other angiogenesis-dependent
CC diseases include Osler-Webber syndrome, myocardial angiogenesis, plaque
CC neovascularisation, telangiectasia, haemophilic joints and wound
CC granulation. In addition, the peptides can be used as birth control
CC agents.
XX
XX SQ Sequence 184 AA;
XX
XX Query Match 100.0%; Score 66; DB 22; Length 184;
XX Best Local Similarity 100.0%; Pred. No. 0.00041;
XX Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 QPVLHLVALNTPL 13
XX DB 7 QPVLHLVALNTPL 19
XX
XX
XX RESULT 15.
XX ABG31793
XX ID ABG31793 standard; Protein; 184 AA.
XX AC ABG31793;
XX
XX DT 05-DEC-2002 (first entry)
XX
XX DE Human endostatin polypeptide.
XX
XX KW Human, endostatin; tumour; cancer; metastasis; cytostatic;
XX antiangiogenic.
XX
XX OS Homo sapiens.
XX
XX PN WO200268457-A2.
XX PD 06-SEP-2002.
XX
XX FF 27-FEB-2002; 2002WO-IT00119.
XX
XX PR 27-FEB-2001; 2001IT-MI00394.
XX
XX (UWMI-) UNIV MILANO.
XX
XX Chillemi F, Vicentini LMT, Francescato P;
XX
XX WPI; 2002-698655/75.
XX
XX New peptide useful for the preparation of medicaments with

```

PT antiangiogenic activity that may be used in treating tumours or  
PT metastases, comprises a sequence corresponding to fragments of human  
PT endostatin -

XX  
XX  
PS Disclosure: Fig 1; 24pp; English.

XX  
CC The invention relates to peptide comprising 20-50 amino acids with  
CC sequences corresponding to the human endostatin polypeptide sequence, its  
CC salt or non-toxic derivative. The peptides are useful in the preparation  
CC of medicaments with antiangiogenic activity which may be useful in  
CC treating tumours or metastases. This sequence represents a human  
CC endostatin polypeptide.

XX  
SQ Sequence 184 AA;

Query Match 100.0%; Score 66; DB 23; Length 184;  
Best Local Similarity 100.0%; Pred. No. 0.00041;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
| | | | | | | | | | | | | | |  
Db 7 QPVLHLVALNTPL 19

Search completed: January 26, 2004, 19:06:22  
Job time : 74 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 26, 2004, 19:04:52 ; Search time 28 Seconds  
(without alignments)  
19.644 Million cell updates/sec

Title: US-09-766-412-30

Perfect score: 66

Sequence: 1 QPVLHLVALNTPL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pcp.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pcp.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pcp.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pcp.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS COMB.pcp.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	100.0	13	3	US-09-385-442-30
2	66	100.0	20	2	US-08-740-168A-1
3	66	100.0	20	3	US-09-349-429-1
4	66	100.0	20	4	US-09-315-689-1
5	66	100.0	20	4	US-09-174-282-1
6	66	100.0	185	3	US-08-985-526-36
7	66	100.0	191	4	US-09-561-500-13
8	66	100.0	191	4	US-09-561-108-13
9	66	100.0	191	4	US-09-561-526-13
10	66	100.0	191	4	US-09-561-499-13
11	66	100.0	195	1	US-08-159-784-2
12	62	93.9	178	4	US-09-315-689-5
13	62	93.9	182	4	US-09-561-500-14
14	62	93.9	182	4	US-09-561-108-14
15	62	93.9	182	4	US-09-315-689-3
16	62	93.9	182	4	US-09-561-526-14
17	62	93.9	182	4	US-09-561-499-14
18	62	93.9	183	3	US-09-206-059-2
19	44	66.7	191	1	US-08-159-784-3
20	41	62.1	955	4	US-09-252-991A-18882
21	40	60.6	413	4	US-09-724-623-87
22	38	57.6	2311	4	US-08-934-386-9
23	37	56.1	187	1	US-08-471-058-17
24	37	56.1	187	3	US-08-471-057-17
25	37	56.1	187	4	US-08-470-865-17
26	37	56.1	191	1	US-08-607-269-28
27	37	56.1	191	5	PCT-US95-04600-28

#### ALIGNMENTS

##### RESULT 1

US-09-385-442-30  
; Sequence 30, Application US/09385442  
; Patent No. 6200954  
; GENERAL INFORMATION:  
; APPLICANT: Ge, Ruowen  
; APPLICANT: Kini, R. Manjunatha  
; TITLE OF INVENTION: Small Peptides Having Potent Anti-Angiogenic Activity  
; FILE REFERENCE: 1781-170P  
; CURRENT APPLICATION NUMBER: US/09/385,442  
; CURRENT FILING DATE: 1999-08-30  
; EARLIER APPLICATION NUMBER: 60/099,313  
; EARLIER FILING DATE: 1999-09-04  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 30  
; LENGTH: 13  
; TYPE: PRT  
; ORGANISM: mammalian  
; FEATURE:  
; OTHER INFORMATION: Endo-2  
US-09-385-442-30

Query Match 100.0%; Score 66; DB 3; Length 13;  
Best Local Similarity 100.0%; Pred. No. 3.6e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13

DB 1 QPVLHLVALNTPL 13

##### RESULT 2

US-08-740-168A-1  
; Sequence 1, Application US/08740168A  
; Patent No. 5854205  
; GENERAL INFORMATION:  
; APPLICANT: O'Reilly, Michael  
; APPLICANT: Folkman, M. Judan  
; TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions  
; TITLE OF INVENTION: and Methods  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jones & Askew, LLP  
; STREET: 191 Peachtree, 37th Floor  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303  
; COMPUTER READABLE FORM:

Sequence 48, Appl  
Sequence 4649, Ap  
Sequence 29546, A  
Sequence 4, Appl  
Sequence 4, Appl  
Sequence 60, Appl  
Sequence 205, App  
Sequence 4748, Ap  
Sequence 19928, A  
Sequence 2, Appl  
Sequence 26149, A  
Sequence 17229, A  
Sequence 19440, A  
Sequence 5, Appl  
Sequence 742, App  
Sequence 1677, Ap  
Sequence 1677, Ap

28 37 56.1 2475 3 US-09-413-814-48  
29 36 54.5 131 4 US-09-107-532A-4649  
30 36 54.5 334 4 US-09-252-991A-29546  
31 36 54.5 342 3 US-09-330-611-4  
32 36 54.5 511 3 US-08-991-677-4  
33 35 53.0 84 4 US-09-300-008B-60  
34 35 53.0 108 4 US-09-732-210-205  
35 35 53.0 222 4 US-09-134-001C-4748  
36 35 53.0 287 4 US-09-252-991A-19928  
37 35 53.0 378 4 US-09-186-489-2  
38 35 53.0 378 4 US-10-043-665B-2  
39 35 53.0 445 4 US-09-252-991A-26149  
40 35 53.0 672 4 US-09-252-991A-17229  
41 35 53.0 951 4 US-09-252-991A-19440  
42 35 53.0 1090 3 US-09-085-199B-5  
43 34 51.5 172 4 US-09-198-452A-742  
44 34 51.5 250 4 US-09-702-705-1677  
45 34 51.5 250 4 US-09-736-457-1677

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/740,168A  
FILING DATE: 22-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0223  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Murine  
TISSUE TYPE: Collagen  
US-08-740-168A-1

Query Match 100.0%; Score 66; DB 2; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.8e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
DB 7 QPVLHLVALNTPL 19

RESULT 3  
US-09-349-429-1  
Sequence 1, Application US/09349429  
Patent No. 6174861  
GENERAL INFORMATION:  
APPLICANT: O'Reilly, Michael  
APPLICANT: Folkman, M. Judah  
TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions  
TITLE OF INVENTION: and Methods  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew, LLP  
STREET: 191 Peachtree, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/349,429  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/740,168  
FILING DATE: 22-OCT-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren, William L.  
REGISTRATION NUMBER: 36,714  
REFERENCE/DOCKET NUMBER: 05213-0223  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 404-818-3700  
TELEFAX: 404-818-3799  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: internal  
ORIGINAL SOURCE:  
ORGANISM: Murine  
TISSUE TYPE: Collagen  
US-09-349-429-1

Query Match 100.0%; Score 66; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.8e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
DB 7 QPVLHLVALNTPL 19

RESULT 4  
US-09-315-689-1  
Sequence 1, Application US/09315689  
Patent No. 6346510  
GENERAL INFORMATION:  
APPLICANT: Folkman, Judah  
APPLICANT: O'Reilly, Michael  
TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions  
FILE REFERENCE: 05213-0229  
CURRENT APPLICATION NUMBER: US/09/315,689  
CURRENT FILING DATE: 1999-05-20  
NUMBER OF SEQ ID NOS: 6  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1  
LENGTH: 20  
TYPE: PRT  
ORGANISM: murine  
US-09-315-689-1

Query Match 100.0%; Score 66; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 5.8e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
DB 7 QPVLHLVALNTPL 19

RESULT 5  
US-09-174-282-1  
Sequence 1, Application US/09174282  
Patent No. 6544758  
GENERAL INFORMATION:  
APPLICANT: O'Reilly, Michael  
APPLICANT: Folkman, M. Judah  
TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions  
TITLE OF INVENTION: and Methods  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Jones & Askew, LLP  
STREET: 191 Peachtree, 37th Floor  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/174,282
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/740,168
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0223
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Murine
; TISSUE TYPE: Collagen
; US-09-174-282-1

Query Match 100.0%; Score 66; DB 4; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.8e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
DB 7 QPVLHLVALNTPL 19

RESULT 6
US-09-985-526-36
; Sequence 36, Application US/08985526
; Patent No. 6080728
; GENERAL INFORMATION:
; APPLICANT: Mixson, James A
; TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA
; TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Connolly, Bove, Lodge, & Hutz
; STREET: 1220 Market Street, P.O. Box 2207
; CITY: Wilmington
; STATE: Delaware
; COUNTRY: U.S.A.
; ZIP: 19899
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,526
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,845
; FILING DATE: 16-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: McMorow Jr., Robert G
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (302) 658-9141
; TELEFAX: (302) 658-5613
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 185 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-985-526-36

Query Match 100.0%; Score 66; DB 3; Length 185;
Best Local Similarity 100.0%; Pred. No. 0.00064;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
DB 8 QPVLHLVALNTPL 20

RESULT 7
US-09-561-500-13
; Sequence 13, Application US/09561500
; Patent No. 6342219
; GENERAL INFORMATION:
; APPLICANT: Phillip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002500
; CURRENT APPLICATION NUMBER: US/09/561,500
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; US-09-561-500-13

Query Match 100.0%; Score 66; DB 4; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.00067;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
DB 14 QPVLHLVALNTPL 26

RESULT 8
US-09-561-108-13
; Sequence 13, Application US/09561108
; Patent No. 6342221
; GENERAL INFORMATION:
; APPLICANT: Phillip E. Thorpe
; APPLICANT: Rolf A. Brekken
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF
; FILE REFERENCE: 4001.002584
; CURRENT APPLICATION NUMBER: US/09/561,108
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/131,432
; PRIOR FILING DATE: 1999-04-28
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
; US-09-561-108-13

```

Query Match 100.0%; Score 66; DB 4; Length 191;  
Best Local Similarity 100.0%; Pred. No. 0.00067; Indels 0; Gaps 0;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
Db 14 QPVLHLVALNTPL 26

RESULT 9  
US-09-561-526-13  
; Sequence 13, Application US/09561526  
; Patent No. 6416758  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brekken  
; TITLE OF INVENTION: ANTIBODY CONJUGATE KITS FOR SELECTIVELY INHIBITING VEGF  
; FILE REFERENCE: 4001.002586  
; CURRENT APPLICATION NUMBER: US/09/561,526  
; CURRENT FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/131,432  
; PRIOR FILING DATE: 1999-04-28  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-561-526-13

Query Match 100.0%; Score 66; DB 4; Length 191;  
Best Local Similarity 100.0%; Pred. No. 0.00067;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
Db 14 QPVLHLVALNTPL 26

RESULT 10  
US-09-561-499-13  
; Sequence 13, Application US/09561499  
; Patent No. 6524583  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brekken  
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF  
; FILE REFERENCE: 4001.002582  
; CURRENT APPLICATION NUMBER: US/09/561,499  
; CURRENT FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/131,432  
; PRIOR FILING DATE: 1999-04-28  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-561-499-13

Query Match 100.0%; Score 66; DB 4; Length 191;  
Best Local Similarity 100.0%; Pred. No. 0.00067;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
Db 14 QPVLHLVALNTPL 26

RESULT 11  
US-08-159-784-2  
; Sequence 2, Application US/08159784  
; Patent No. 5643783  
; GENERAL INFORMATION:  
; APPLICANT: Bjorn R. Olsen  
; TITLE OF INVENTION: NOVEL COLLAGEN AND USES THEREOF  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/159,784  
; FILING DATE: December 1, 1993  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: John F. Freeman  
; REGISTRATION NUMBER: 29,066  
; REFERENCE/DOCKET NUMBER: 00246/170001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 195  
; TYPE: amino acid  
; STRANDEDNESS: N/A  
; TOPOLOGY: N/A  
US-08-159-784-2

Query Match 100.0%; Score 66; DB 1; Length 195;  
Best Local Similarity 100.0%; Pred. No. 0.00068;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
Db 18 QPVLHLVALNTPL 30

RESULT 12  
US-09-315-689-5  
; Sequence 5, Application US/09315689  
; Patent No. 6346510  
; GENERAL INFORMATION:  
; APPLICANT: Polkman, Judah  
; APPLICANT: O'Reilly, Michael  
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions  
; FILE REFERENCE: 05213-0229  
; CURRENT APPLICATION NUMBER: US/09/315,689  
; CURRENT FILING DATE: 1999-05-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 178  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-315-689-5

Query Match 93.9%; Score 62; DB 4; Length 178;  
Best Local Similarity 92.3%; Pred. No. 0.0029;

Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 QPVLHLVALNTPL 13  
Db 3 QPVLHLVALNSPL 15

RESULT 13  
US-09-561-500-14  
; Sequence 14, Application US/09561500  
; Patent No. 6342219  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brekken  
; TITLE OF INVENTION: ANTIBODY COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF  
; FILE REFERENCE: 4001.002500  
; CURRENT APPLICATION NUMBER: US/09/561.500  
; CURRENT FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/131,432  
; PRIOR FILING DATE: 1999-04-28  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 182  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
; OTHER INFORMATION: PEPTIDE  
US-09-561-500-14

Query Match 93.9%; Score 62; DB 4; Length 182;  
Best Local Similarity 92.3%; Pred. No. 0.003;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPVLHLVALNTPL 13  
Db 7 QPVLHLVALNSPL 19

RESULT 14  
US-09-561-108-14  
; Sequence 14, Application US/09561108  
; Patent No. 6342221  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brekken  
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY INHIBITING VEGF  
; FILE REFERENCE: 4001.002584  
; CURRENT APPLICATION NUMBER: US/09/561.108  
; CURRENT FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/131,432  
; PRIOR FILING DATE: 1999-04-28  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 182  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
; OTHER INFORMATION: PEPTIDE  
US-09-561-108-14

Query Match 93.9%; Score 62; DB 4; Length 182;  
Best Local Similarity 92.3%; Pred. No. 0.003;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPVLHLVALNTPL 13  
Db 7 QPVLHLVALNSPL 19

RESULT 15  
US-09-315-689-3  
; Sequence 3, Application US/09315689  
; Patent No. 6346510  
; GENERAL INFORMATION:  
; APPLICANT: Folkman, Judah  
; APPLICANT: O'Reilly, Michael  
; TITLE OF INVENTION: Therapeutic Antiangiogenic Endostatin Compositions  
; FILE REFERENCE: 05213-0229  
; CURRENT APPLICATION NUMBER: US/09/315,689  
; CURRENT FILING DATE: 1999-05-20  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 182  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-315-689-3  
Query Match 93.9%; Score 62; DB 4; Length 182;  
Best Local Similarity 92.3%; Pred. No. 0.003;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 QPVLHLVALNTPL 13  
Db 7 QPVLHLVALNSPL 19  
Search completed: January 26, 2004, 19:10:12  
Job time : 30 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 26, 2004, 19:06:31 ; Search time 367 Seconds  
(without alignments)  
7,323 Million cell updates/sec

Title: US-09-766-412-30

Perfect score: 66

Sequence: 1 QPVLHLVALNTPL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 777136 seqs, 206736638 residues

Total number of hits satisfying chosen parameters: 777136

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	100.0	13	10	US-09-766-412-30
2	66	100.0	20	9	US-09-873-676-115
3	66	100.0	20	10	US-09-405-499-1
4	66	100.0	20	10	US-09-174-516-1
5	66	100.0	20	10	US-09-154-302-1
6	66	100.0	20	12	US-10-351-284-1
7	66	100.0	20	15	US-10-131-241-43
8	66	100.0	20	15	US-10-232-316-1
9	66	100.0	20	15	US-10-042-347-1
10	66	100.0	184	12	US-10-292-418-18
11	66	100.0	184	15	US-10-131-241-46
12	66	100.0	185	14	US-10-036-869-36
13	66	100.0	191	10	US-09-998-831-13
14	66	100.0	191	12	US-10-373-561-13
15	66	100.0	207	12	US-10-422-934-71

16	66	100.0	207	14	US-10-080-797-3	Sequence 3, Appli
17	62	93.9	178	15	US-10-131-241-60	Sequence 60, Appli
18	62	93.9	178	15	US-10-042-347-5	Sequence 5, Appli
19	62	93.9	179	15	US-10-131-241-57	Sequence 57, Appli
20	62	93.9	180	15	US-10-131-241-47	Sequence 47, Appli
21	62	93.9	180	15	US-10-131-241-54	Sequence 54, Appli
22	62	93.9	181	15	US-10-131-241-55	Sequence 55, Appli
23	62	93.9	182	10	US-09-998-831-14	Sequence 14, Appli
24	62	93.9	182	12	US-10-373-561-14	Sequence 54, Appli
25	62	93.9	182	15	US-10-131-241-54	Sequence 54, Appli
26	62	93.9	182	15	US-10-042-347-3	Sequence 3, Appli
27	62	93.9	183	9	US-09-873-676-2	Sequence 2, Appli
28	62	93.9	183	12	US-10-292-418-4	Sequence 4, Appli
29	62	93.9	183	14	US-10-080-797-1	Sequence 1, Appli
30	62	93.9	183	15	US-10-131-241-52	Sequence 52, Appli
31	62	93.9	682	12	US-10-264-049-3010	Sequence 3010, Ap
32	62	93.9	684	11	US-09-961-403-5	Sequence 5, Appli
33	62	93.9	1516	12	US-10-431-642-3	Sequence 3, Appli
34	62	93.9	1516	15	US-10-060-036-166	Sequence 166, App
35	58	87.9	184	12	US-10-292-418-35	Sequence 35, Appli
36	58	87.9	184	15	US-10-131-241-49	Sequence 49, Appli
37	48	72.7	184	12	US-09-938-391-4	Sequence 4, Appli
38	48	72.7	230	12	US-09-938-391-2	Sequence 2, Appli
39	43	65.2	1867	11	US-09-824-574-5	Sequence 5, Appli
40	40	60.6	413	12	US-10-288-930-87	Sequence 87, Appli
41	40	60.6	556	15	US-10-128-714-3458	Sequence 3458, Ap
42	40	60.6	556	15	US-10-128-714-3584	Sequence 3584, Ap
43	40	60.6	576	15	US-10-128-714-8458	Sequence 8458, Ap
44	40	60.6	576	15	US-10-128-714-8584	Sequence 8584, Ap
45	39	59.1	181	15	US-10-149-819-14	Sequence 14, Appli

#### ALIGNMENTS

RESULT 1  
US-09-766-412-30  
; Sequence 30, Application US/09766412  
; Patent No. US20020103129A1  
; GENERAL INFORMATION:  
; APPLICANT: GE. Ruwen et al.  
; TITLE OF INVENTION: SMALL PEPTIDES HAVING ANTI-ANGIOGENIC AND ENDOTHELIAL CELL INHIBI  
; TITLE OF INVENTION: ACTIVITY  
; FILE REFERENCE: 1781-0215P  
; CURRENT APPLICATION NUMBER: US/09/766,412  
; CURRENT FILING DATE: 2001-01-11  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 30  
; LENGTH: 13  
; TYPE: PRT  
; ORGANISM: Mammalian  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Endo-2  
US-09-766-412-30

Query Match 100.0%; Score 66; DB 10; Length 13;  
Best Local Similarity 100.0%; Pred. No. 4.6e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
DB 1 QPVLHLVALNTPL 13

RESULT 2  
US-09-873-676-115  
; Sequence 115, Application US/09873676  
; Patent No. US20020077289A1  
; GENERAL INFORMATION:  
; APPLICANT: MacDonald, Nicholas J.  
; APPLICANT: Sim, Kim L.

```
; TITLE OF INVENTION: Angiostatin and Endostatin Binding Proteins and Methods of Use
; FILE REFERENCE: 05213-0378 (43170-259333)
; CURRENT APPLICATION NUMBER: US/09/873.676
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: US 60/209,065
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: US 60/289,387
; PRIOR FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 115
; LENGTH: 20
; TYPE: PRT
; ORGANISM: mammalian
US-09-873-676-115

Query Match      100.0%; Score 66; DB 9; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 QPVLHLVALNTPL 13
Db      7 QPVLHLVALNTPL 19

RESULT 3
US-09-405-499-1
; Sequence 1, Application US/09405499
; Patent No. US20020123458A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Endostatin Protein and Fragments Thereof
; FILE REFERENCE: 05213-0640
; CURRENT APPLICATION NUMBER: US/09/405,499
; CURRENT FILING DATE: 1999-09-23
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Murinae gen. sp.
US-09-405-499-1

Query Match      100.0%; Score 66; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 QPVLHLVALNTPL 13
Db      7 QPVLHLVALNTPL 19

RESULT 4
US-09-174-516-1
; Sequence 1, Application US/09174516A
; Patent No. US20020127595A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael S.
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Methods of Detecting Endostatin Protein
; FILE REFERENCE: 05213-0227
; CURRENT APPLICATION NUMBER: US/09/174,516A
; CURRENT FILING DATE: 1998-10-16
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Murinae gen. sp.
US-09-174-516-1

Query Match      100.0%; Score 66; DB 10; Length 20;
```

```
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 QPVLHLVALNTPL 13
Db      7 QPVLHLVALNTPL 19

RESULT 5
US-09-154-302-1
; Sequence 1, Application US/09154302
; Patent No. US20020155987A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions
; TITLE OF INVENTION: and Methods
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew, LLP
; STREET: 191 Peachtree, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,302
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/740,168
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0223
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Murine
; TISSUE TYPE: Collagen
US-09-154-302-1

Query Match      100.0%; Score 66; DB 10; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 QPVLHLVALNTPL 13
Db      7 QPVLHLVALNTPL 19

RESULT 6
US-10-351-284-1
; Sequence 1, Application US/10351284
; Publication No. US20030219426A1
; GENERAL INFORMATION:
```

```

; APPLICANT: O'Reilly, Michael S.
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions and Methods
; FILE REFERENCE: 05213-3051 (43170-282623)
; CURRENT APPLICATION NUMBER: US/10/351,284
; CURRENT FILING DATE: 2003-01-24
; PRIOR APPLICATION NUMBER: US 60/005,835
; PRIOR FILING DATE: 1995-10-23
; PRIOR APPLICATION NUMBER: US 60/023,070
; PRIOR FILING DATE: 1996-08-02
; PRIOR APPLICATION NUMBER: US 60/026,263
; PRIOR FILING DATE: 1996-09-17
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Murinae gen. sp.
US-10-351-284-1

Query Match 100.0%; Score 66; DB 12; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
Db 7 QPVLHLVALNTPL 19

RESULT 7
US-10-131-241-43
; Sequence 43, Application US/10131241
; Publication No. US20030012792A1
; GENERAL INFORMATION:
; APPLICANT: Holaday, John W.
; APPLICANT: Fortier, Anne H.
; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Proliferation
; FILE REFERENCE: 05213-0344 43170-271565
; CURRENT APPLICATION NUMBER: US/10/131,241
; CURRENT FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: US 09/413,049
; PRIOR FILING DATE: 1999-10-06
; PRIOR APPLICATION NUMBER: US 09/316,802
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: US 60/086,586
; PRIOR FILING DATE: 1998-05-22
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 43
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Murinae sp.
US-10-131-241-43

Query Match 100.0%; Score 66; DB 15; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
Db 7 QPVLHLVALNTPL 19

RESULT 8
US-10-232-316-1
; Sequence 1, Application US/10232316
; Publication No. US20030087393A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Therapeutic Antiangiogenic Compositions and Methods

```

```

; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jones & Askew, LLP
; STREET: 191 Peachtree, 37th Floor
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/10/232,316
; FILING DATE: 03-Sep-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/740,168A
; FILING DATE: 22-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren, William L.
; REGISTRATION NUMBER: 36,714
; REFERENCE/DOCKET NUMBER: 05213-0223
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404-818-3700
; TELEFAX: 404-818-3799
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Murine
; TISSUE TYPE: Collagen
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-232-316-1

Query Match 100.0%; Score 66; DB 15; Length 20;
Best Local Similarity 100.0%; Pred. No. 7.3e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
Db 7 QPVLHLVALNTPL 19

RESULT 9
US-10-042-347-1
; Sequence 1, Application US/10042347
; Publication No. US20030114370A1
; GENERAL INFORMATION:
; APPLICANT: O'Reilly, Michael S.
; APPLICANT: Folkman, M. Judah
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Endostatin Protein and Peptide Fragments
; FILE REFERENCE: 05213-0880 (43170-249874)
; CURRENT APPLICATION NUMBER: US/10/042,347
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: US 09/315,689
; PRIOR FILING DATE: 1999-05-20
; PRIOR APPLICATION NUMBER: US 60/106,343
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: US 09/154,302
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: US 08/740,168
; PRIOR FILING DATE: 1996-10-22
; PRIOR APPLICATION NUMBER: US 60/005,835

```



;; PRIOR FILING DATE: 1995-10-23  
;; PRIOR APPLICATION NUMBER: US 60/023,070  
;; PRIOR FILING DATE: 1996-08-32  
;; PRIOR APPLICATION NUMBER: US 60/026,263  
;; PRIOR FILING DATE: 1996-09-17  
;; NUMBER OF SEQ ID NOS: 6  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 1  
;; LENGTH: 20  
;; TYPE: PRT  
;; ORGANISM: Murinae sp.  
US-10-042-347-1

Query Match 100.0%; Score 66; DB 15; Length 20;  
Best Local Similarity 100.0%; Pred. No. 7.3e-05;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
Db 7 QPVLHLVALNTPL 19

RESULT 10  
US-10-292-418-18  
;; Sequence 18, Application US/10292418  
;; Publication No. US20030139365A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Li, Yue  
;; APPLICANT: Gillies, Stephen D  
;; TITLE OF INVENTION: Expression and Export of Angiogenesis Inhibitors as  
;; FILE REFERENCE: LEX-006C1  
;; CURRENT APPLICATION NUMBER: US/10/292,418  
;; PRIOR FILING DATE: 2002-11-12  
;; PRIOR APPLICATION NUMBER: 09/383,315  
;; PRIOR FILING DATE: 1999-08-25  
;; PRIOR APPLICATION NUMBER: US 60/097,883  
;; PRIOR FILING DATE: 1998-08-25  
;; NUMBER OF SEQ ID NOS: 54  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 18  
;; LENGTH: 184  
;; TYPE: PRT  
;; ORGANISM: Mus musculus  
US-10-292-418-18

Query Match 100.0%; Score 66; DB 12; Length 184;  
Best Local Similarity 100.0%; Pred. No. 0.00078;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
Db 7 QPVLHLVALNTPL 19

RESULT 11  
US-10-131-241-46  
;; Sequence 46, Application US/10131241  
;; Publication No. US20030012792A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Holaday, John W.  
;; APPLICANT: Fortier, Anne H.  
;; TITLE OF INVENTION: Compositions and Methods for Inhibiting Endothelial Cell Prolifer  
;; TITLE OF INVENTION: and Regulating Angiogenesis Using Cancer Markers  
;; FILE REFERENCE: 05213-0344 43170-271565  
;; CURRENT APPLICATION NUMBER: US/10/131,241  
;; CURRENT FILING DATE: 2002-07-22  
;; PRIOR APPLICATION NUMBER: US 09/413,049  
;; PRIOR FILING DATE: 1999-10-06  
;; PRIOR APPLICATION NUMBER: US 09/316,802  
;; PRIOR FILING DATE: 1999-05-21  
;; PRIOR APPLICATION NUMBER: US 60/086,586

;; PRIOR FILING DATE: 1998-05-22  
;; NUMBER OF SEQ ID NOS: 65  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 46  
;; LENGTH: 184  
;; TYPE: PRT  
;; ORGANISM: Murinae sp.  
US-10-131-241-46

Query Match 100.0%; Score 66; DB 15; Length 184;  
Best Local Similarity 100.0%; Pred. No. 0.00078;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
Db 7 QPVLHLVALNTPL 19

RESULT 12  
US-10-036-869-36  
;; Sequence 36, Application US/10036869  
;; Publication No. US20020151516A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Mixson, James A  
;; TITLE OF INVENTION: CARRIER:DNA COMPLEXES CONTAINING DNA  
;; ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE  
;; THERAPY  
;; NUMBER OF SEQUENCES: 43  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Connolly, Bove, Lodge, & Hutz  
;; STREET: 1220 Market Street, P.O. Box 2207  
;; CITY: Wilmington  
;; STATE: Delaware  
;; COUNTRY: U.S.A.  
;; ZIP: 19899  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/10/036,869  
;; FILING DATE: 29-NO. US20020151516A1-2001  
;; CLASSIFICATION: <Unknown>  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/985,526  
;; FILING DATE: <Unknown>  
;; APPLICATION NUMBER: US 08/608,845  
;; FILING DATE: 16-JUL-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: McMorow Jr., Robert G  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (302) 658-9141  
;; TELEFAX: (302) 658-5613  
;; INFORMATION FOR SEQ ID NO: 36:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 185 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; SEQUENCE DESCRIPTION: SEQ ID NO: 36:  
US-10-036-869-36

Query Match 100.0%; Score 66; DB 14; Length 185;  
Best Local Similarity 100.0%; Pred. No. 0.00079;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13  
Db 8 QPVLHLVALNTPL 20

RESULT 13  
US-09-998-831-13

; Sequence 13, Application US/09998831  
; Patent No. US2002011915A1  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brekken  
; TITLE OF INVENTION: ANTIBODY CONJUGATE COMPOSITIONS FOR SELECTIVELY  
; TITLE OF INVENTION: INHIBITING VEGF  
; FILE REFERENCE: 4001.002584  
; CURRENT APPLICATION NUMBER: US/09/998,831  
; CURRENT FILING DATE: 2003-11-30  
; PRIOR APPLICATION NUMBER: 09/561,108  
; PRIOR FILING DATE: 2000-04-28  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-998-831-13

Query Match 100.0%; Score 66; DB 10; Length 191;  
Best Local Similarity 100.0%; Pred. No. 0.00082;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPVLHLVALNTPL 13  
|||  
Db 14 QPVLHLVALNTPL 26

RESULT 14  
US-10-373-561-13  
; Sequence 13, Application US/10373561  
; Publication No. US20030175276A1  
; GENERAL INFORMATION:  
; APPLICANT: Philip E. Thorpe  
; APPLICANT: Rolf A. Brekken  
; TITLE OF INVENTION: ANTIBODY METHODS FOR SELECTIVELY INHIBITING VEGF  
; FILE REFERENCE: 4001.002582  
; CURRENT APPLICATION NUMBER: US/10/373,561  
; CURRENT FILING DATE: 2003-02-24  
; PRIOR APPLICATION NUMBER: US/09/561,499  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 60/131,432  
; PRIOR FILING DATE: 1999-04-28  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 13  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-10-373-561-13

Query Match 100.0%; Score 66; DB 12; Length 191;  
Best Local Similarity 100.0%; Pred. No. 0.00082;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPVLHLVALNTPL 13  
|||  
Db 14 QPVLHLVALNTPL 26

RESULT 15  
US-10-422-934-71  
; Sequence 71, Application US/10422934  
; Publication No. US20030186841A1  
; GENERAL INFORMATION:  
; APPLICANT: Barbas, Carlos F., III  
; APPLICANT: Kadan, Michael  
; APPLICANT: Beerli, Roger

; TITLE OF INVENTION: LIGAND ACTIVATED TRANSCRIPTIONAL REGULATOR PROTEINS  
; FILE REFERENCE: 22908-1227C  
; CURRENT APPLICATION NUMBER: US/10/422,934  
; CURRENT FILING DATE: 2003-04-23  
; PRIOR APPLICATION NUMBER: 09/586,625  
; PRIOR FILING DATE: 2000-06-02  
; PRIOR APPLICATION NUMBER: 09/433,042  
; PRIOR FILING DATE: 1999-10-25  
; NUMBER OF SEQ ID NOS: 92  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 71  
; LENGTH: 207  
; TYPE: PRT  
; ORGANISM: Muridae  
US-10-422-934-71

Query Match 100.0%; Score 66; DB 12; Length 207;  
Best Local Similarity 100.0%; Pred. No. 0.00089;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QPVLHLVALNTPL 13  
|||  
Db 30 QPVLHLVALNTPL 42

Search completed: January 26, 2004, 19:16:33  
Job time : 368 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 26, 2004, 19:04:50 ; Search time 38 Seconds  
(without alignments)  
32,900 Million cell updates/sec

Title: US-09-766-412-30

Perfect score: 66

Sequence: 1 QPVLHLVALNPL 13

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 76:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	100.0	1315	2 A56101	collagen alpha 1(XVIII)
2	66	100.0	1774	2 B56101	collagen alpha 1(XVIII)
3	62	93.9	684	2 A3019	collagen alpha 1(XVIII)
4	44	66.7	1388	2 A3317	collagen alpha 1(XVIII)
5	43	65.2	1867	2 S2275	MOT1 protein - yea
6	40	60.6	176	2 C70602	hypothetical prote
7	40	60.6	342	2 A1045	conserved hypothet
8	39	59.1	190	2 D75319	3-methyladenine gl
9	39	59.1	333	2 I40791	acetoin dehydrogen
10	39	59.1	425	2 A80796	NADH2 dehydrogenas
11	39	59.1	533	2 G96541	probable cytochrom
12	38	57.6	103	2 T50934	hypothetical prote
13	38	57.6	160	2 E72631	hypothetical prote
14	38	57.6	258	2 T25958	hypothetical prote
15	38	57.6	336	2 T44958	hypothetical prote
16	38	57.6	617	2 T15104	hypothetical prote
17	38	57.6	2311	2 T06161	acetyl-CoA carboxy
18	38	57.6	3206	1 GNVSVU	genome polypeptid
19	37	56.1	102	2 S0397	probable membrane
20	37	56.1	191	1 QBE4	BHRF1 protein - hu
21	37	56.1	519	2 A36542	probable cytochrom
22	37	56.1	728	1 TVHUSK	transforming prote
23	36	54.5	72	1 PNBPF6	p13 protein - phag
24	36	54.5	298	2 T2264	hypothetical prote
25	36	54.5	323	2 T48425	lipase-like protei
26	36	54.5	331	2 A3534	probable C4-dicarb
27	36	54.5	342	1 S56374	hypothetical 38.7K
28	36	54.5	342	2 G31269	hypothetical prote
29	36	54.5	342	2 E86110	hypothetical prote

30 36 54.5 358 2 T15069  
31 36 54.5 370 2 H72748  
32 36 54.5 548 2 G82986  
33 36 54.5 651 2 AG0977  
34 36 54.5 656 2 C91186  
35 36 54.5 656 2 B86033  
36 36 54.5 1162 2 T37889  
37 35 53.0 111 2 F83574  
38 35 53.0 190 1 FWECK  
39 35 53.0 190 2 B85569  
40 35 53.0 190 2 D90719  
41 35 53.0 193 2 A81149  
42 35 53.0 133 2 H81874  
43 35 53.0 135 2 A50587  
44 35 53.0 222 2 H71360  
45 35 53.0 222 2 B72729

## ALIGNMENTS

RESULT 1  
A56101  
collagen alpha 1(XVIII) chain precursor, short splice form - mouse  
N;Contains: endostatin  
C;Species: Mus musculus (house mouse)  
C;Date: 03-Oct-1995 #sequence\_revision 08-May-1998 #text\_change 31-Mar-2000  
C;Accession: A56101; A58371; S72450; S65595  
R;Rehn, M.; Pihlajaniemi, T.  
J. Biol. Chem. 270, 4705-4711, 1995  
A;Title: Identification of three N-terminal ends of type XVIII collagen chains and tissue  
tif homologous to rat and Drosophila frizzled proteins.  
A;Reference number: A56101; MUID:95181468; PMID:7876242  
A;Accession: A56101  
A;Molecule type: mRNA  
A;Residues: 1-103 <REH1>  
A;Cross-references: GB:U11636; NID:9618427; PIDN:AAC52178.1; PID:G618428  
R;Rehn, M.; Pihlajaniemi, T.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4234-4238, 1994  
A;Title: Alpha1(XVIII) a collagen chain with frequent interruptions in the collagenous  
A;Reference number: A58371; MUID:94240112; PMID:8183894  
A;Accession: A58371  
A;Molecule type: mRNA  
A;Residues: 1-928 <REH2>  
A;Cross-references: GB:I16898; NID:9404754; PIDN:AAA37434.1; PID:G553894  
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
submitted to the EMBL Data Library, August 1993  
A;Reference number: S72450  
A;Accession: S72450  
A;Molecule type: mRNA  
A;Residues: 28-687, 'F', 736-751, 'R', 753-1315 <ORW>  
A;Cross-references: EMBL:L22545; NID:G348968; PIDN:AAA19787.1; PID:G511298  
R;Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
A;Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-  
A;Reference number: A58370; MUID:94240111; PMID:8183893  
A;Accession: S65595  
A;Molecule type: mRNA  
A;Residues: 28-1315 <OHS>  
A;Cross-references: EMBL:L22545  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit (C  
lated and subsequently O-glycosylated.  
C;Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in peri  
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un  
ay be useful in treating solid tumors.  
C;Genetics:  
A;Gene: MGI:Col18a1  
A;Cross-references: MGI:71175  
A;Map position: 10:41.0  
C;Superfamily: unassigned collagens  
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc  
F;1-25/Domain: signal sequence #status predicted <SIG>  
F;24-235/Region: thrombospondin amino-terminal similarity

F.26-1315/Product: collagen alpha 1(XVIII) chain, short splice form #status predicted <M>  
 F.321-353/Domain: collagenous #status predicted <CO1>  
 F.364-437/Domain: collagenous #status predicted <CO2>  
 F.462-583/Domain: collagenous #status predicted <CO3>  
 F.607-689/Domain: collagenous #status predicted <CO4>  
 F.704-745/Domain: collagenous #status predicted <CO5>  
 F.759-831/Domain: collagenous #status predicted <CO6>  
 F.843-874/Domain: collagenous #status predicted <CO7>  
 F.887-910/Domain: collagenous #status predicted <CO8>  
 F.892-894/Domain: collagenous #status predicted <CO9>  
 F.918-969/Domain: collagenous #status predicted <CO10>  
 F.983-1000/Domain: collagenous #status predicted <CO11>  
 F.1132-1315/Product: endostatin #status predicted <EST>  
 F.1139-1315/Region: multiplexin collagen carboxyl-terminal similarity  
 F.126-488/Binding site: carboxydrate (Ser) (covalent) #status predicted  
 F.172-228/Disulfide bonds: #status predicted  
 F.240-454,1257/Binding site: carboxydrate (Ser) (covalent) #status predicted  
 F.451,454,1257/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 66; DB 2; Length 1315;  
 Best Local Similarity 100.0%; Pred. No. 0.00055; Indels 0; Gaps 0;  
 Matches 13; Conservative 0; Mismatches 0;

QY 1 QPVHLVALNTPPL 13  
 |||||  
 DB 1138 QPVHLVALNTPPL 1150

RESULT 2  
 B56101  
 N:Contains: collagen alpha 1(XVIII) chain precursor, long splice form - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 03-Oct-1995 #sequence revision 08-May-1998 #text change 31-Mar-2000  
 A:Accession: B56101; C56101; S72450; S65595; PNO675; A54072; A58816  
 R:Rehn, M.; Pihlajaniemi, T.  
 J. Biol. Chem. 270, 4705-4711, 1995  
 A:Title: Identification of three N-terminal ends of type XVIII collagen chains and tissue homologous to rat and Drosophila frizzled proteins.  
 A:Reference number: A56101; MUID:95181468; PMID:7876242  
 A:Accession: B56101  
 A:Molecule type: mRNA  
 A:Residues: 1-362 <REH1>  
 A:Cross-references: GB:U11637; NID:G618430; PID:AC52179.1; PID:G618430  
 A:Experimental source: splice form clone PE17.24  
 A:Accession: C56101  
 A:Molecule type: mRNA  
 A:Residues: 1-239,487-562 <REH2>  
 A:Cross-references: GB:U11637; NID:G618429  
 A:Experimental source: splice form clones PB8.1, PB19, PB15.2  
 R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, B.R.  
 submitted to the EMBL Data Library, August 1993  
 A:Reference number: S72450  
 A:Accession: S72450  
 A:Molecule type: mRNA  
 A:Residues: 487-1146 'L', 1148-1193 'F', 1195-1210 'R', 1212-1512 'L', 1514-1522 'F', 1524-1616 'R', 1618-1633 'V', 1685-1774 <OH2>  
 A:Cross-references: EMBL:L22545; NID:G34968; PID:AAA19787.1; PID:G511298  
 R:Oh, S.P.; Kamagata, Y.; Muragaki, Y.; Timmons, S.; Ooshima, A.; Olsen, A.B.R.  
 Proc. Natl. Acad. Sci. U.S.A. 91, 4229-4233, 1994  
 A:Title: Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa  
 A:Reference number: A58370; MUID:94240111; PMID:8183893  
 A:Accession: S65595  
 A:Molecule type: mRNA  
 A:Residues: 487-1512 'L', 1514-1522 'F', 1524-1693 'V', 1685-1774 <OH2>  
 A:Cross-references: EMBL:L22545  
 R:Abé, N.; Muragaki, Y.; Yoshioka, H.; Inoue, H.; Ninomiya, Y.  
 Biochem. Biophys. Res. Commun. 196, 576-582, 1993  
 A:Title: Identification of a novel collagen chain represented by extensive interruptions  
 A:Reference number: PNO675; MUID:94C59075; PMID:8240330  
 A:Accession: PNO675  
 A:Molecule type: mRNA  
 A:Residues: 635-1774 <AB>  
 R:Rehn, M.; Hintikka, E.; Pihlajaniemi, T.

J. Biol. Chem. 269, 13929-13935, 1994  
 A:Title: Primary structure of the alpha chain of mouse type XVIII collagen, partial str collagen chain.  
 A:Reference number: A54072; MUID:94245707; PMID:8188673  
 A:Accession: A54072  
 A:Molecule type: DNA; mRNA  
 A:Residues: 1293-1403, R', 1405-1774 <REH3>  
 A:Cross-references: GB:U03714; NID:9487733; PID:9487734  
 R:O'Reilly, M.S.; Boehm, T.; Shing, Y.; Fukai, N.; Vasios, G.; Lane, W.S.; Flynn, E.; B Cell 88, 277-285, 1997  
 A:Title: Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.  
 A:Reference number: A58816; MUID:97160848; PMID:9008168  
 A:Accession: A58816  
 A:Molecule type: Protein  
 A:Residues: 1591-1610 <RE>  
 A:Experimental source: hemangioendothelium cells  
 A:Note: Inhibits endothelial cell proliferation  
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit lated and subsequently O-glycosylated.  
 C:Comment: The different splice forms of collagen alpha 1(XVIII) may be involved in per C:Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of ur ay be useful in treating solid tumors.  
 C:Genetics:  
 A:Gene: MGI:Coll18a1  
 A:Cross-references: MGI:71175  
 A:Map position: 10:41.0  
 A:Introns: 1295/3; 1310/1; 1331/1; 1345/3; 1388/3; 1437/1; 1461/3; 1505/3; 1516/3; 1599/3  
 A:Note: the list of introns is incomplete  
 C:Superfamily: unassigned collagens  
 C:Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc F.1-1774/Product: collagen alpha 1(XVIII) chain precursor, long splice form #status prec F.1-239,487-1774/Product: collagen alpha 1(XVII) chain precursor, medium splice form # F.1-24/Domain: signal sequence #status predicted <SIG>  
 F.361-486/Region: frizzled similarity  
 F.786-812/Domain: collagenous #status predicted <CO01>  
 F.823-896/Domain: collagenous #status predicted <CO02>  
 F.921-1042/Domain: collagenous #status predicted <CO03>  
 F.1066-1148/Domain: collagenous #status predicted <CO04>  
 F.1163-1204/Domain: collagenous #status predicted <CO05>  
 F.1218-1290/Domain: collagenous #status predicted <CO06>  
 F.1301-1333/Domain: collagenous #status predicted <CO07>  
 F.1346-1369/Domain: collagenous #status predicted <CO08>  
 F.1351-1353/Region: cell attachment (R-G-D) motif  
 F.1377-1428/Domain: collagenous #status predicted <CO09>  
 F.1442-1459/Domain: collagenous #status predicted <CO10>  
 F.1591-1774/Product: endostatin #status predicted <EST>  
 F.1598-1774/Region: multiplexin collagen carboxyl-terminal similarity  
 F.354,361,947/Binding site: carboxydrate (Asn) (covalent) #status predicted  
 F.599,704,1716/Binding site: carboxydrate (Ser) (covalent) #status predicted  
 F.910,913,1053/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 66; DB 2; Length 1774;  
 Best Local Similarity 100.0%; Pred. No. 0.00076; Indels 0; Gaps 0;  
 Matches 13; Conservative 0; Mismatches 0;

QY 1 QPVHLVALNTPPL 13  
 |||||  
 DB 1597 QPVHLVALNTPPL 1609

RESULT 3  
 A53019  
 collagen alpha 1(XVIII) chain - human (fragment)  
 N:Contains: endostatin  
 C:Species: Homo sapiens (man)  
 C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 31-Mar-2000  
 C:Accession: A53019  
 R:Oh, S.P.; Warman, M.L.; Seidlin, M.F.; Cheng, S.D.; Knoll, J.H.M.; Timmons, S.; Olsen, Genomics 19, 494-499, 1994  
 A:Title: Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localiza A:Reference number: A53019; MUID:94245237; PMID:8188291  
 A:Accession: A53019  
 A:Molecule type: mRNA

A;Residues: 1-594 <OHA>  
A;Cross-references: GB:L22548; NID:9348908; PIDN:AAA51864.1; PID:9562794  
A;Note: the cited accession number, L25548, is not in Genbank release 103  
A;Note: in the authors' translation, 482-Gly is not shown, residues 483-490 are shifted  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
labeled and subsequently O-glycosylated.  
C;Comment: Different splice forms of collagen alpha 1(XVIII) may be involved in perivasc  
C;Comment: Endostatin is released from collagen alpha 1(XVIII) chain by the action of un  
ay be useful in treating solid tumors.  
C;Genetics:  
A;Gene: GDB:COL18A1  
A;Cross-references: GDB:138752; ONTM:120328  
A;Map position: 21q22.3-21q22.3  
C;Superfamily: unassigned collagens  
C;Keywords: alternative splicing; angiogenesis inhibitor; chondroitin sulfate proteoglyc  
F;1-684/Product: collagen alpha 1(XVIII) chain (fragment) #status predicted <MAT>  
F;1-59/Domain: collagenous (fragment) #status predicted <CO4>  
F;74-115/Domain: collagenous #status predicted <CO5>  
F;129-201/Domain: collagenous #status predicted <CO6>  
F;212-244/Domain: collagenous #status predicted <CO7>  
F;257-278/Domain: collagenous #status predicted <CO8>  
F;262-264/Region: cell attachment (R-G-D) motif  
F;285-340/Domain: collagenous #status predicted <CO9>  
F;354-371/Domain: collagenous #status predicted <CO10>  
F;503-684/Product: endostatin #status predicted <EST>  
F;503-684/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 93.9%; Score 62; DB 2; Length 684;  
Best Local Similarity 92.3%; Pred. No. 0.0015;  
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTP 13  
|||||  
DB 508 QPVLHLVALNSPL 520  
|||||

RESULT 4  
A53317  
collagen alpha 1(XV) chain precursor - human  
N;Alternate names: procollagen alpha 1(XV) chain  
C;Species: Homo sapiens (man)  
C;Date: 07-Jul-1995 #sequence revision 07-Jul-1995 #text\_change 31-Mar-2000  
C;Accession: A53317; A53146; S28778  
R;Keywords: S.; Heinecke, P.; Rehn, M.; Honkanen, N.; Myers, J.C.; Pihlajaniemi, T.  
J. Biol. Chem. 269, 4773-4779, 1994  
A;Title: Primary structure of the alpha1 chain of human type XV collagen and exon-intron  
A;Reference number: A53317; MUID:94148920; PMID:8106446  
A;Accession: A53317  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1388 <XIV>  
A;Cross-references: GB:L25280  
A;Note: nucleotide sequence and conceptual translation not complete  
R;Muragaki, Y.; Abe, N.; Ninomiya, Y.; Olsen, B.R.; Ooshima, A.  
J. Biol. Chem. 269, 4042-4046, 1994  
A;Title: The human alpha1(XV) collagen chain contains a large amino-terminal non-triple  
A;Reference number: A53146; MUID:94140817; PMID:8307960  
A;Accession: A53146  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-9, 'S', 11-48, 'V', 50-94, 'A', 96-149, 'A', 151-203, 'V', 205-408, 'A', 410-569 <MUR>  
A;Cross-references: GB:D21230; NID:9415605; PIDN:BA04762.1; PID:d1005294; PID:9460703  
R;Myers, J.C.; Kivirikko, S.; Gordon, M.K.; Dion, A.S.; Pihlajaniemi, T.  
Proc. Natl. Acad. Sci. U.S.A. 89, 10144-10148, 1992  
A;Title: Identification of a previously unknown human collagen chain, alpha1(XV), charac  
A;Reference number: S28778; MUID:93065196; PMID:1279671  
A;Accession: S28778  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 544-640, 'P', 642-811, 'P', 813-1252 <MYE>  
C;Genetics:  
A;Gene: GDB:COL15A1  
A;Cross-references: GDB:132578; ONTM:120325

A;Map position: 9q21-9q22  
A;Superfamily: unassigned collagens  
F;1-22/Domain: signal sequence #status predicted <SIG>  
F;13-138/Region: multiplexin collagen carboxyl-terminal similarity  
F;1216-1388/Region: multiplexin collagen carboxyl-terminal similarity

Query Match 66.7%; Score 44; DB 2; Length 1388;  
Best Local Similarity 66.7%; Pred. No. 8.3;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTP 12  
|||||  
DB 1215 KPALHLAALNMP 1226  
|||||

RESULT 5  
S22775  
MOT1 protein - yeast (Saccharomyces cerevisiae)  
N;Alternate names: protein LPP4c; protein YPL082c  
C;Species: Saccharomyces cerevisiae  
C;Date: 17-Apr-1993 #sequence revision 17-Apr-1993 #text\_change 21-Jul-2000  
C;Accession: S22775; S61106  
R;Davis, J.L.; Kunisawa, R.; Thorne, J.  
Mol. Cell. Biol. 12, 1879-1892, 1992  
A;Title: A presumptive helicase (MOT1 gene product) affects gene expression and is requir  
A;Reference number: S22775; MUID:92195335; PMID:1312673  
A;Accession: S22775  
A;Molecule type: DNA  
A;Residues: 1-1867 <DAV>  
A;Cross-references: EMBL:M83224; NID:G171964; PIDN:AAA34786.1; PID:G171965  
R;Hall, J.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.; Vo, D.H.; War  
submitted to the EMBL Data Library, August 1995  
A;Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.  
A;Reference number: S59677  
A;Accession: S61106  
A;Molecule type: DNA  
A;Residues: 1-1867 <HAL>  
A;Cross-references: EMBL:U41849; NID:G1147608; PID:G1147612; MIPS:YPL082C  
C;Genetics:  
A;Gene: SGD:MOT1  
A;Cross-references: SGD:S0006003; MIPS:YPL082C  
A;Map position: 16L  
C;Keywords: DNA binding; nucleus; transmembrane protein  
F;700-716/Domain: transmembrane #status predicted <TM1>  
F;1038-1054/Domain: transmembrane #status predicted <TM2>  
F;1186-1202/Domain: transmembrane #status predicted <TM3>

Query Match 65.2%; Score 43; DB 2; Length 1867;  
Best Local Similarity 69.2%; Pred. No. 18;  
Matches 9; Conservative 2; Mismatches 0; Indels 2; Gaps 1;

QY 1 QPVLHLVALNTP 13  
|||||  
DB 637 QPILHL--LNTPV 647  
|||||

RESULT 6  
C70602  
hypothetical protein Rv1000 - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence revision 17-Jul-1998 #text\_change 20-Jun-2000  
C;Accession: C70602  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: C70602  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-176 <COL>

A;Cross-references: GB:294752; GB:AL123456; NID:G3261731; PIDN:CAB08143.1; PID:G2052135  
A;Experimental source: strain H37Rv  
C;Genetics:  
C;Superfamily: Mycobacterium tuberculosis hypothetical protein Rv1000

Query Match 60.6%; Score 40; DB 2; Length 176;  
Best Local Similarity 63.6%; Pred. No. 4.8;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PVLHLVALNTP 12  
| | | | |  
Db 103 PWHLTALGSP 113  
| | | | |

RESULT 7  
A;Accession: AH1045  
A;Title: conserved hypothetical protein yJek [imported] - Salmonella enterica subsp. enterica serovar typhi  
C;Species: Salmonella enterica subsp. enterica serovar typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AH1045  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar typhi  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AH1045  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-342 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD06813.1; PID:G16505463; GSPDB:GN00176  
C;Genetics:  
C;Superfamily: conserved hypothetical protein yodo

Query Match 60.6%; Score 40; DB 2; Length 342;  
Best Local Similarity 77.8%; Pred. No. 10;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 5 HVALNTP 13  
| | | | |  
Db 3 HIVTUNTP 11  
| | | | |

RESULT 8  
D75319  
3-methyladenine glycosidase - Deinococcus radiodurans (strain R1)  
C;Species: Deinococcus radiodurans  
C;Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 15-Jun-2001  
C;Accession: D75319  
R;White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma, S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A;Reference number: A75250; MUID:20036896; PMID:10567266  
A;Accession: D75319  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-190 <WHI>  
A;Cross-references: GB:AE002043; GB:AE000513; NID:G6459859; PIDN:AAFL1623.1; PID:G645986  
A;Experimental source: strain R1  
C;Genetics:  
A;Gene: DR2074  
C;Superfamily: Bacillus subtilis DNA-3-methyladenine glycosidase homolog yxjJ

Query Match 59.1%; Score 39; DB 2; Length 190;  
Best Local Similarity 66.7%; Pred. No. 8.1;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PVLHLVALNTP 13  
| | | | |  
Db 136 PELHLLAPETPL 147  
| | | | |

RESULT 9  
I40791  
acetoin dehydrogenase (TPP-dependent) (EC 1.-.-.-) beta chain - Clostridium magnum  
C;Species: Clostridium magnum  
C;Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 11-Jun-1999  
C;Accession: I40791  
R;Kruger, N.; Oppermann, F.B.; Lorenzl, H.; Steinbuechel, A.  
J. Bacteriol. 176, 3614-3630, 1994  
A;Title: Biochemical and molecular characterization of the Clostridium magnum acetoin dehydrogenase  
A;Reference number: I40789; MUID:94266715; PMID:8206840  
A;Accession: I40791  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-333 <XRU>  
A;Cross-references: GB:I31844; NID:G473324; PIDN:AAA21745.1; PID:G47327  
C;Superfamily: pyruvate dehydrogenase (lipoamide) beta chain  
C;Keywords: oxidoreductase

Query Match 59.1%; Score 39; DB 2; Length 333;  
Best Local Similarity 58.3%; Pred. No. 15;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PVLHLVALNTP 13  
| | | | |  
Db 297 PVRIGALNTP 308  
| | | | |

RESULT 10  
AB0796  
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) - Salmonella enterica subsp. enterica serovar typhi  
C;Species: Salmonella enterica subsp. enterica serovar typhi  
A;Note: this species has also been called Salmonella typhi  
C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
C;Accession: AB0796  
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.  
Nature 413, 848-852, 2001  
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar typhi  
A;Reference number: AB0502; MUID:21534947; PMID:11677608  
A;Accession: AB0796  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-425 <PAR>  
A;Cross-references: GB:AL513382; PIDN:CAD07548.1; PID:G16503540; GSPDB:GN00176  
C;Genetics:  
A;Gene: STY2546  
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 2  
C;Keywords: oxidoreductase

Query Match 59.1%; Score 39; DB 2; Length 425;  
Best Local Similarity 61.5%; Pred. No. 20;  
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 QPVLHLVALNTP 13  
| | | | |  
Db 412 QPLISLVQATPL 424  
| | | | |

RESULT 11  
G96541  
probable cytochrome P450 [imported] - Arabidopsis thaliana  
C;Species: Arabidopsis thaliana (mouse-ear cress)  
C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 20-Apr-2001  
C;Accession: G96541  
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: G96541  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-533 <STO>  
A:Cross-references: GB:AE005173; NID:gi11054632; PIDN:AAG27877.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: FltJ6.6  
A:Map position: 1  
A:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology  
C:Keywords: heme; iron; metalloprotein  
F:457/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 59.1%; Score 39; DB 2; Length 533;  
Best Local Similarity 58.3%; Pred. No. 25;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 PVLHLVALNTPL 13  
Db 77 PLLHLAFNIPI 88

RESULT 12  
T50954  
hypothetical protein B24P7.90 [imported] - Neurospora crassa  
C:Species: Neurospora crassa  
C:Date: 21-Jul-2000 #sequence\_revision 21-Jul-2000 #text\_change 08-Sep-2000  
A:Accession: T50954  
R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
submitted to the Protein Sequence Database, July 2000  
A:Reference number: Z25286  
A:Accession: T50954  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-103 <SCH>  
A:Cross-references: EMBL:AL398890; GSPDB:GN00116; NCSP:B24P7.90  
A:Experimental source: BAC clone B24P7; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B24P7.90  
A:Map position: 6  
A:Superfamily: Neurospora crassa hypothetical protein B24P7.90

Query Match 57.6%; Score 38; DB 2; Length 103;  
Best Local Similarity 58.3%; Pred. No. 6.4;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 QPVLHLVALNTPL 12  
Db 70 QPVLHLQLESP 81

RESULT 13  
E72631  
hypothetical protein APE1508 - Aeropyrum pernix (strain K1)  
C:Species: Aeropyrum pernix  
C:Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
A:Accession: E72631  
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Takah  
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K  
DNA Res. 6, 83-101, 1999  
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr  
A:Reference number: A72450; MUID:99310339; PMID:10382966  
A:Accession: E72631  
A:Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-160 <KAW>  
A:Cross-references: DDBJ:AP000061; NID:gs104821; PIDN:BAA80507.1; PID:dl044293; PID:gs101  
A:Experimental source: strain K1  
C:Genetics:  
A:Gene: APE1508

Query Match 57.6%; Score 38; DB 2; Length 160;  
Best Local Similarity 50.0%; Pred. No. 10;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PVLHLVALNTPL 13  
Db 56 PVLHLPLDIPI 67

RESULT 14  
T25958  
hypothetical protein ZC204.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
A:Accession: T25958  
R:Wamsley, P.; Kramer, J.  
submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans cosmid ZC204.  
A:Reference number: Z20116  
A:Accession: T25958  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-258 <WAM>  
A:Cross-references: EMBL:U80839; PIDN:AA837912.1; GSPDB:GN00020; CESP:ZC204.3  
A:Experimental source: strain Bristol N2; clone ZC204  
C:Genetics:  
A:Gene: CESP:ZC204.3  
A:Map position: 2  
A:Introns: 5/1; 123/3; 167/1

Query Match 57.6%; Score 38; DB 2; Length 258;  
Best Local Similarity 80.0%; Pred. No. 18;  
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 LHLVALNTPL 13  
Db 85 LYLVALETPL 94

RESULT 15  
T44958  
hypothetical protein [imported] - Natronomonas pharaonis  
C:Species: Natronomonas pharaonis  
C:Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jan-2000  
A:Accession: T44958  
R:Mattar, S.; Souquet, M.; Heinrich, H.J.; Engelhard, M.  
submitted to the EMBL Data Library, August 1996  
A:Description: The first fully resolved primary structure of an archaeal succinate-dehydi  
A:Reference number: Z22881  
A:Accession: T44958  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-336 <MAT>  
A:Cross-references: EMBL:Y07709; PIDN:CAA68978.1  
A:Experimental source: strain SP1 /28

Query Match 57.6%; Score 38; DB 2; Length 336;  
Best Local Similarity 63.6%; Pred. No. 24;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 PVLHLVALNTPL 12  
Db 312 PLCHLVALDEP 322

Search completed: January 26, 2004, 19:08:54

Job time : 42 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 26, 2004, 19:04:50 ; Search time 23 Seconds  
(without alignments)

26,580 Million cell updates/sec

Title: US-09-766-412-30  
Perfect score: 66  
Sequence: 1 QPVLHVALNTP 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	100.0	1527	1 CAIH_MOUSE	P39061 mus musculus
2	62	93.9	1516	1 CAIH_HUMAN	P39060 homo sapien
3	44	66.7	1388	1 CAIE_HUMAN	P39059 homo sapien
4	43	65.2	1867	1 MOT1_YEAST	P32333 saccharomyc
5	40	60.6	217	1 ATP6_RHOD	Q8b686 rhopalosiph
6	39	59.1	190	1 MGH_DEIRA	Q8rsq0 deinococcus
7	38	57.6	3206	1 POLG_PSNV	P29152 p genome po
8	37	56.1	102	1 YN21_YEAST	P40211 saccharomyc
9	37	56.1	191	1 EAR_EBV	P01182 Epstein-bar
10	37	56.1	348	1 SKI_MOUSE	Q80598 mus musculus
11	37	56.1	728	1 SKI_HUMAN	P12755 homo sapien
12	36	54.5	72	1 VP13_BPH6	P11130 bacterioph
13	36	54.5	323	1 UCP4_HUMAN	Q95847 homo sapien
14	36	54.5	342	1 YUEK_ECOLI	P39280 escherichia
15	36	54.5	370	1 EGSA_AERPE	Q9yer2 aeropyrum p
16	36	54.5	749	1 NIBL_MOUSE	Q8r1f1 mus musculus
17	35	53.0	190	1 ATKC_ECO57	Q8x9G0 escherichia
18	35	53.0	190	1 ATKC_ECOL6	Q8f1v5 escherichia
19	35	53.0	190	1 ATKC_ECOLI	P03961 escherichia
20	35	53.0	193	1 LOLE_NEIMA	P57023 neisseria m
21	35	53.0	193	1 LOLE_NEIMB	P57024 reisseria m
22	35	53.0	194	1 ATKC_SALT1	Q8z86c salmonella
23	35	53.0	194	1 ATKC_SALT2	Q8z9G0 salmonella
24	35	53.0	222	1 GIDB_TREPA	Q66106 treponema p
25	35	53.0	329	1 ANFK_CLOHU	Q46084 clostridium
26	35	53.0	332	1 G3P_EHYIN	P26988 phytophthor
27	35	53.0	348	1 ARC2_DEIRA	Q8rvq3 deinococcus
28	35	53.0	387	1 CYB_SCHPO	P05501 schizosacch
29	35	53.0	397	1 LIPG_BOVIN	Q29459 bos taurus
30	35	53.0	462	1 ANFK_AZOVI	P16267 azotobacter
31	35	53.0	618	1 ORC2_DROME	Q24168 drosophila
32	35	53.0	623	1 HCVE_EURCA	P02242 eurytelma c
33	35	53.0	956	1 RRPO_SBNV	P21405 southern be

#### RESULT 1

CAIH\_MOUSE ID CAIH\_MOUSE STANDARD; PRT; 1527 AA.  
AC P39061; Q61437; Q62002;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(XVIII) chain precursor (Contains: Endostatin).  
GN COL18A1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
[1]\_RN  
RP SEQUENCE FROM N.A. (ISOFORM SHORT).  
RC STRAIN=BALB/c; TISSUE=Liver;  
RX MEDLINE=94245707; PubMed=8188673;  
RA Rehn M.V., Hintikka E., Pihlajaniemi T.;  
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen, partial structure of the corresponding gene, and comparison of the alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen chain.";  
RT J. Biol. Chem. 269:13929-13935(1994).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).  
RA Rehn M., Hintikka E., Pihlajaniemi T.;  
RT "Characterization of the mouse gene for the alpha-1 chain of type XVIII collagen (COL18A1) reveals that the three variant N-terminal polypeptide forms are transcribed from two widely separated promoters.";  
RT Promoters.";  
RT Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 213-1140 FROM N.A. (ISOFORM SHORT).  
RX MEDLINE=94240112; PubMed=8183894;  
RA Rehn M.V., Pihlajaniemi T.;  
RT "Alpha 1(XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen.";  
RT Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).  
RN [4]  
RP SEQUENCE OF 240-1527 FROM N.A.  
RX TISSUE=Liver;  
RN MEDLINE=94240111; PubMed=8183893;  
RA Oh S.P., Kamagata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.;  
RT "Isolation and sequencing of cDNAs for proteins with multiple domains of Gly-Xaa-Yaa repeats identify a distinct family of collagenous proteins.";  
RT Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).  
RN [5]  
RP CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.  
RX MEDLINE=97160848; PubMed=9008168;  
RA O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vasios G., Lane W.S., Flynn E., Birkhead J.R., Olsen B.R., Folkman J.;  
RT "Endostatin: an endogenous inhibitor of angiogenesis and tumor growth.";  
RT Cell 88:277-285(1997).

P12689 saccharomyc  
O83273 treponema p  
P02948 rhodobacter  
P02949 rhodobacter  
P07116 escherichia  
Q8u9e0 agrobacteri  
Q9v1y5 pyrococcus  
Q00275 apis mellif  
Q9b1y9 homo sapien  
P30138 escherichia  
P33832 variola vir  
P07193 xenopus lae

#### ALIGNMENTS



Query Match 100.0%; Score 66; DB 1; Length 1527;  
Best Local Similarity 100.0%; Pred. No. 0.00038;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHVALNTP 13  
DB 1350 QPVLHVALNTP 1362

RESULT 2  
ID CALH HUMAN STANDARD; PRT; 1516 AA.  
AC P39060; Q9UK38; Q9Y6Q7; Q9Y6Q8;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 16-OCT-2001 (Rel. 46, Last sequence update)  
DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].  
GN COL18A1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98164096; PubMed=9503365;  
RA Saarela J., Ilikarppa R., Rehn M., Purmonen S., Pihlajaniemi T.;  
RT "Complete primary structure of two variant forms of human type XVIII  
collagen and tissue-specific differences in the expression of the  
corresponding transcripts.";  
RL Matrix Biol. 16:319-328(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20289799; PubMed=10830953;  
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,  
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,  
RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,  
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,  
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,  
RA Rosenthal A., Kudoh J., Shibuya K., Kawaesaki K., Asakawa S.,  
RA Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E.,  
RA Minoshima S., Shimizu N., Nordiek G., Hornischer K., Brandt P.,  
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,  
RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,  
RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,  
RA Lehrach H., Reinhardt R., Vaspo M.-L.;  
RT "The DNA sequence of human chromosome 21.";  
RL Nature 405:311-319(2000).  
RN [3]  
RP SEQUENCE OF 834-1516 FROM N.A.  
RX MEDLINE=94245237; PubMed=8188291;  
RA Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,  
RA Olsen B.R.;  
RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen  
and localization of the alpha 1(XVIII) collagen gene to mouse  
chromosome 10 and human chromosome 21.";  
RL Genomics 19:494-499(1994).  
RN [4]  
RP SEQUENCE OF 1334-1516 FROM N.A.  
RC TISSUE=Placenta;  
RA Zhi-Yong H., Bao L., Wei-Jie Z., Xiang-Fu W.;  
RT "Cloning and expression of human endostatin gene in Escherichia  
coli.";  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP INVOLVEMENT IN KNOBLOCH SYNDROME.  
RX MEDLINE=20400145; PubMed=10942434;  
RA Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,  
RA Passos-Bueno M.R.;  
RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis  
and tumor growth, plays a critical role in the maintenance of retinal  
structure and in neural tube closure.";  
RT Hum. Mol. Genet. 9:2051-2058(2000).

[6]  
VARIANT ASN-1437.  
RX MEDLINE=21518361; PubMed=11606364;  
RA Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,  
RA Zorick T., Soares F., Camargo A., Oliveira E.S., di Loreto C.,  
RA Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;  
RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes  
for the development of prostatic adenocarcinoma.";  
RL Cancer Res. 61:7375-7378(2001).  
CC -!- FUNCTION: COL18A1 PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE  
RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.  
CC -!- FUNCTION: ENDOSTATIN POTENTIALLY INHIBITS ENDOTHELIAL CELL  
PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY  
BINDING TO THE HEPARAN SULPHATE PROTEOGLYCAN INVOLVED IN GROWTH  
FACTOR SIGNALLING.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=Long; Synonyms=NC-493;  
CC IsoId=P39060-1; Sequence=Displayed;  
CC Name=Short; Synonyms=NC1-303;  
CC IsoId=P39060-2; Sequence=VSP\_001155, VSP\_001156;  
CC -!- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS  
IN LIVER, LUNG AND KIDNEY.  
CC -!- PTM: Prolines at the third position of the tripeptide repeating  
unit (G-X-Y) are hydroxylated in some or all of the chains.  
CC -!- POLYMORPHISM: There is an association between a polymorphism in  
position 1437 and prostate cancer. Heterozygous Asn-1437  
individuals have a 2.5 times increased chance of developing  
prostate cancer as compared with homozygous Asp-1437 individuals.  
CC -!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KNO)  
[MIM:267750]; an autosomal recessive disorder defined by the  
occurrence of high myopia, vitreoretinal degeneration with retinal  
detachment, macular abnormalities and occipital encephalocele.  
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
INTERRUPTED HELICES (FACIT) FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF018081; AAC39658.1; -;  
CC EMBL; AF018082; AAC39659.1; -;  
CC EMBL; AL163302; CAB90482.1; -;  
CC EMBL; L22548; AAA51864.1; -;  
CC EMBL; AF184060; AAF01310.1; ALT\_INIT.  
CC PDB; 1BNL; 02-BEC-98.  
CC GlycoSuiteDB; P39060;  
CC Genew; HGNC:2195; COL18A1.  
CC MIM; 120328; -;  
CC MIM; 267750; -;  
CC GO; GO:0005581; C:collagen; TAS.  
CC GO; GO:0008181; F:tumor suppressor; TAS.  
CC GO; GO:0007397; P:histogenesis and organogenesis; TAS.  
CC GO; GO:0008285; P:negative regulation of cell proliferation; TAS.  
CC GO; GO:0007048; P:oncogenesis; TAS.  
CC GO; GO:0007601; P:vision; TAS.  
CC InterPro; IPR000087; Collagen.  
CC InterPro; IPR001791; Laminin\_G.  
CC InterPro; IPR003129; TSEN.  
CC Pfam; PF01391; Collagen; 7.  
CC Pfam; PF02210; TSPN; 1.  
CC ProDom; PDC00007; Clg\_helix; 1.  
CC SMART; SM00282; LamG; 1.  
CC SMART; SM00210; TSPN; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;  
KW Polymorphism; 3D-structure. POTENTIAL.  
FT SIGNAL 1 23  
FT CHAIN 24 1516 COLLAGEN ALPHA 1 (XVIII) CHAIN.

FT	CHAIN	1334	1516	ENDOSTATIN.	01-FEB-1995 (Rel. 31, Created)
FT	DOMAIN	24	516	NONHELICAL REGION 1 (NC1).	01-FEB-1995 (Rel. 31, Last sequence update)
FT	DOMAIN	517	550	TRIPLE-HELICAL REGION 1 (COL1).	28-FEB-2003 (Rel. 41, Last annotation update)
FT	DOMAIN	551	560	NONHELICAL REGION 2 (NC2).	Collagen alpha 1(XV) chain precursor.
FT	DOMAIN	561	640	TRIPLE-HELICAL REGION 2 (COL2).	COL15A1.
FT	DOMAIN	641	664	NONHELICAL REGION 3 (NC3).	Homo sapiens (Human).
FT	DOMAIN	665	786	TRIPLE-HELICAL REGION 3 (COL3).	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
FT	DOMAIN	787	809	TRIPLE-HELICAL REGION 4 (NC4).	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
FT	DOMAIN	810	892	TRIPLE-HELICAL REGION 4 (COL4).	NCBI_TaxID=9606;
FT	DOMAIN	893	906	NONHELICAL REGION 5 (NC5).	(1)
FT	DOMAIN	907	948	TRIPLE-HELICAL REGION 5 (COL5).	SEQUENCE FROM N.A.
FT	DOMAIN	949	961	NONHELICAL REGION 6 (NC6).	TISSUE=Umbilical cord;
FT	DOMAIN	962	1034	TRIPLE-HELICAL REGION 6 (COL6).	MEDLINE=94148920; PubMed=8106446;
FT	DOMAIN	1035	1044	NONHELICAL REGION 7 (NC7).	Kivirikko S., Heinamaki P., Rehn M.V., Honkanen N., Myers J.C.,
FT	DOMAIN	1045	1077	TRIPLE-HELICAL REGION 7 (COL7).	Pihlajaniemi T.;
FT	DOMAIN	1078	1089	NONHELICAL REGION 8 (NC8).	"Primary structure of the alpha 1 chain of human type XV collagen and
FT	DOMAIN	1090	1111	TRIPLE-HELICAL REGION 8 (COL8).	exon-intron organization in the 3' region of the corresponding
FT	DOMAIN	1112	1118	NONHELICAL REGION 9 (NC9).	Gene.";
FT	DOMAIN	1119	1173	TRIPLE-HELICAL REGION 9 (COL9).	J. Biol. Chem. 269:4773-4779(1994).
FT	DOMAIN	1174	1186	NONHELICAL REGION 10 (NC10).	(2)
FT	DOMAIN	1187	1204	TRIPLE-HELICAL REGION 10 (COL10).	SEQUENCE OF 1-569 FROM N.A.
FT	DOMAIN	1205	1516	NONHELICAL REGION 11 (NC11).	TISSUE=Placenta;
FT	CARBOHYD	68	68	N-LINKED (GLCNAC. . .) (POTENTIAL).	MEDLINE=94140817; PubMed=8307960;
FT	CARBOHYD	129	129	N-LINKED (GLCNAC. . .) (POTENTIAL).	Muragaki Y., Abe N., Ninomiya Y., Olsen B.R., Ooshima A.;
FT	CARBOHYD	154	164	N-LINKED (GLCNAC. . .) (POTENTIAL).	"The human alpha 1(XV) collagen chain contains a large amino-terminal
FT	CARBOHYD	691	691	N-LINKED (GLCNAC. . .) (POTENTIAL).	non-triple helical domain with a tandem repeat structure and homology
FT	CARBOHYD	1329	1329	O-LINKED (GALNAC. . .) (POTENTIAL).	to alpha 1(XVIII) collagen.";
FT	CARBOHYD	1366	1506	BY SIMILARITY.	J. Biol. Chem. 269:4042-4046(1994).
FT	DISULFID	1468	1498	CELL ATTACHMENT SITE (POTENTIAL).	(3)
FT	SITE	1095	1097	Missing (in isoform Short).	SEQUENCE OF 544-1252 FROM N.A.
FT	VARSPLIC	1	180	FTTid=VSP 001155.	MEDLINE=9306196; PubMed=1279671;
FT	VARSPLIC	181	215	HTTAGGLPAPTPSPRLGPRWAPLCPSPVPPSS -> MA	Myers J.C., Kivirikko S., Gordon M.K., Dion A.S., Pihlajaniemi T.;
FT	VARIANT	1437	1437	PRCPWPERRRLDVLAPLVLLGVNRAAEP (in	"Identification of a previously unknown human collagen chain, alpha
FT	VARIANT	1437	1437	isoform Short).	1(XV), characterized by extensive interruptions in the triple-helical
FT	VARIANT	1437	1437	D -> N (increased risk of developing	region.";
FT	VARIANT	1437	1437	prostate cancer).	Proc. Natl. Acad. Sci. U.S.A. 89:10144-10148(1992).
FT	VARIANT	1437	1437	FTTid=VAR 012709.	-!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN INTERNAL ORGANS
FT	VARIANT	1437	1437	F -> S (IN REF. 2).	SUCH AS ADRENAL GLAND, PANCREAS AND KIDNEY.
FT	VARIANT	1437	1437	I -> V (IN REF. 2).	-!- PTM: Prolines at the third position of the tripeptide repeating
FT	VARIANT	1437	1437	V -> L (IN REF. 3).	unit (G-X-Y) are hydroxylated in some or all of the chains.
FT	VARIANT	1437	1437	P -> R (IN REF. 3).	-!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
FT	VARIANT	1437	1437	P -> L (IN REF. 3).	INTERRUPTED HELICES (FACIT) FAMILY.
FT	VARIANT	1437	1437	P -> A (IN REF. 3).	This SWISS-PROT entry is copyright. It is produced through a collaboration
FT	VARIANT	1437	1437	A -> P (IN REF. 3).	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
FT	VARIANT	1437	1437	L -> K (IN REF. 3).	the European Bioinformatics Institute. There are no restrictions on its
FT	VARIANT	1437	1437	P -> G (IN REF. 3).	use by non-profit institutions as long as its content is in no way
FT	VARIANT	1437	1437	L -> Q (IN REF. 3).	modified and this statement is not removed. Usage by and for commercial
FT	VARIANT	1437	1437	P -> T (IN REF. 4).	entities requires a license agreement (see http://www.isb-sib.ch/announce/
FT	VARIANT	1437	1437	S -> Y (IN REF. 4).	or send an email to license@isb-sib.ch).
FT	VARIANT	1437	1437	P -> PGSP (IN REF. 2).	EMBL; L25286; AAA58429.1; -
FT	VARIANT	1437	1437	G -> GQ (IN REF. 3).	EMBL; D21230; BAA04762.1; -
FT	VARIANT	1437	1437	R -> G (IN REF. 3).	EMBL; L01697; -; NOT ANNOTATED_CDS.
FT	VARIANT	1437	1437	A -> G (IN REF. 3).	PIR; A53317; A53317.
FT	VARIANT	1437	1437	LR -> CG (IN REF. 3).	HSSP; P39061; 1XOE.
FT	VARIANT	1437	1437	R -> T (IN REF. 4).	MIIM; 120325; -
FT	VARIANT	1437	1437	S -> Y (IN REF. 4).	InterPro; IPR000087; Collagen.
FT	VARIANT	1437	1437	3C70F29A4476E576 CRC64;	InterPro; IPR001791; Laminin G.
FT	VARIANT	1437	14		

```

FT DOMAIN 733 763 NONHELICAL REGION 3 (NC3).
FT DOMAIN 754 798 TRIPLE-HELICAL REGION 3 (COL3).
FT DOMAIN 799 822 TRIPLE-HELICAL REGION 4 (NC4).
FT DOMAIN 823 867 TRIPLE-HELICAL REGION 4 (COL4).
FT DOMAIN 868 878 TRIPLE-HELICAL REGION 5 (NC5).
FT DOMAIN 879 949 TRIPLE-HELICAL REGION 5 (COL5).
FT DOMAIN 950 983 NONHELICAL REGION 6 (NC6).
FT DOMAIN 984 1013 NONHELICAL REGION 6 (COL6).
FT DOMAIN 1014 1027 TRIPLE-HELICAL REGION 7 (NC7).
FT DOMAIN 1028 1045 TRIPLE-HELICAL REGION 7 (COL7).
FT DOMAIN 1046 1052 NONHELICAL REGION 8 (NC8).
FT DOMAIN 1053 1107 TRIPLE-HELICAL REGION 8 (COL8).
FT DOMAIN 1108 1117 NONHELICAL REGION 9 (NC9).
FT DOMAIN 1118 1132 TRIPLE-HELICAL REGION 9 (COL9).
FT DOMAIN 1133 1388 NONHELICAL REGION 10 (NC10).
FT DOMAIN 1389 1555 4 X TANDEM REPEATS.
FT REPEAT 358 408 1.
FT REPEAT 409 459 2.
FT REPEAT 460 509 3.
FT REPEAT 510 555 4.
FT CARBOHYD 366 306 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 344 324 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 687 697 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 814 814 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1046 1046 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 10 10 C -> S (IN REF. 2).
FT CONFLICT 49 49 D -> V (IN REF. 2).
FT CONFLICT 95 95 L -> A (IN REF. 2).
FT CONFLICT 150 150 P -> A (IN REF. 2).
FT CONFLICT 204 204 R -> V (IN REF. 2).
FT CONFLICT 409 409 R -> A (IN REF. 2).
SQ SEQUENCE 1388 AA; 141930 MW; 60822AD925A3093D CRC64;

Query Match 66.7%; Score 44; DB 1; Length 1388;
Best Local Similarity 66.7%; Pred. No. 4.4; Mismatches 1; Indels 3; Gaps 0;
Matches 8; Conservative 1;

QY 1 QPVLHLVALNTP 12
DB 1215 KPALHLAALNMP 1226

RESULT 4
MOTI YEAST
ID MOTI YEAST STANDARD; PRT; 1867 AA.
AC P3233,
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable helicase MOT1.
GN MOT1 OR YPL082C OR LYP4C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92195335; PubMed=1312673;
RA Davis J.L., Kunisawa R., Thorne J.;
RT "A presumptive helicase (MOT1 gene product) affects gene expression and is required for viability in the yeast Saccharomyces cerevisiae."
RT Mol. Cell. Biol. 12:1879-1892(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=S288c / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., Dipaolo T., Dubois E., Duesterhoeft A.,

```

```

DE ATP synthase A chain (EC 3.6.3.14) (Protein 6).
GN MTATP6 OR ATP6.
OS Rhopalosiphum padi (Bird cherry-oat aphid).
OG Mitochondrion.
CC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
CC Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
CC Aphidoidea; Aphididae; Aphidini; Rhopalosiphum.
CC NCBI_TaxID=40932;
RN [1]
RP SEQUENCE FROM N.A.
RA Martinez-Torres D., Buades C., Latorre A., Moya A.;
RT "Molecular systematics of aphids and their primary endosymbionts.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Key component of the proton channel; it may play a
CC direct role in the translocation of protons across the membrane
CC (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC -!- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
CC core - and CF(0) - the membrane proton channel. CF(1) has five
CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
CC has three main subunits: A, B and C (By similarity).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ280673; CAC28069.1;
CC InterPro: IPR000568; ATPSynt_Asub.
CC Pfam: PF00119; ATP-Synt_A; 1.
CC PRINTS: PR00123; ATPASEA.
CC TIGRFS: TIGR01131; ATP synt_6_or_A; 1.
CC PROSITE: PS00449; ATPASEA; 1.
CC Hydrogen ion transport; CF(0); Mitochondrion; Transmembrane.
CC SEQUENCE 217 AA; 25442 MW; 2F3EEEAAC53B89B4 CRC64;
CC
CC Query Match 60.6%; Score 40; DB 1; Length 217;
CC Best Local Similarity 54.5%; Pred. No. 3.3;
CC Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
CC
CC QY 3 VLHLVALNTPPL 13
CC : |||: |||:
CC 127 IAHLIPLNTPFI 137
CC
CC DB
CC
CC RESULT 6
CC 3MGH-DEIRA STANDARD; PRT; 190 AA.
CC AC Q9RSQ0;
CC DT 28-FEB-2003 (Rel. 41, Created)
CC DT 28-FEB-2003 (Rel. 41, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Putative 3-methyladenine DNA glycosylase (EC 3.2.2.-).
CC DR2074.
CC OS Deinococcus radiodurans.
CC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
CC Deinococcaceae; Deinococcus.
CC NCBI_TaxID=1299;
CC [1]
CC RN
CC SEQUENCE FROM N.A.
CC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;
CC RC MEDLINE=2036896; PubMed=10567266;
CC RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
CC Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
CC Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
CC Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
CC Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,

```

```

RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
CC -!- SIMILARITY: Belongs to the DNA glycosylase MPG family.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE002043; AAF11623.1;
CC PIR: D75319; D75319.
CC HSP: P29372; IENK.
CC TIGR: DR2074; -.
CC HAMAP: MF_09527; -.
CC InterPro: IPR003180; PurDNA_glycosylase.
CC Pfam: PF02245; Pur DNA Glyco; 1.
CC ProDom: PD009649; PurDNA_glycosylase; 1.
CC Hypothetical protein; DNA repair; Hydrolase; Complete proteome.
CC KW SEQUENCE 190 AA; 20819 MW; 05264B48A76B6B2 CRC64;
CC
CC Query Match 59.1%; Score 39; DB 1; Length 190;
CC Best Local Similarity 66.7%; Pred. No. 4.4;
CC Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC
CC QY 2 PVLHLVALNTPPL 13
CC : |||: |||:
CC 136 PEHLALAPETPL 147
CC
CC DB
CC
CC RESULT 7
CC POLG_PSBMV STANDARD; PRT; 3206 AA.
CC AC P29152;
CC DT 01-DEC-1992 (Rel. 24, Created)
CC DT 01-DEC-1992 (Rel. 24, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE Genome polyprotein [Contains: N-terminal protein (PI); Helper
CC component proteinase (EC 3.4.22.45) (HC-Pro); Protein P3; 6 kDa
CC protein 1 (6k1); Cytoplasmic inclusion protein (CI); 6 kDa protein 2
CC (6k2); Genome-linked protein (VPG); Nuclear inclusion protein A (NI-A)
CC (NIA) (EC 3.4.22.44) (49 kDa proteinase) (49 kDa-Pro); Nuclear
CC inclusion protein B (NI-B) (NIB) (RNA-directed RNA polymerase)
CC (EC 2.7.7.48); Coat protein (CP)].
CC OS pea seed-borne mosaic virus (strain DPD1).
CC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;
CC Potyvirus.
CC NCBI_TaxID=31736;
CC [1]
CC RN
CC SEQUENCE FROM N.A.
CC MEDLINE=9204431; PubMed=1940858;
CC RA Johansen E., Rasmussen O.F., Heide M., Borkhardt B.;
CC RT "The complete nucleotide sequence of pea seed-borne mosaic virus
CC RNA.";
CC RL J. Gen. Virol. 72:2625-2632(1991).
CC -!- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID
CC TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.
CC -!- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT
CC MAY BE INVOLVED IN REPLICATION.
CC -!- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.
CC -!- CATALYTIC ACTIVITY: Hydrolyzes glutamyl bonds, and activity is
CC further restricted by preferences for the amino acids in P6 - P1,
CC that vary with the species of potyvirus, e.g. Glu-Xaa-Xaa-Tyr-Xaa-
CC Gin+(Ser or Gly) for the enzyme from tobacco etch virus. The
CC natural substrate is the viral polyprotein, but other proteins and
CC oligopeptides containing the appropriate consensus sequence are
CC also cleaved.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +

```

```
CC (RNA) (N).
CC -!- CATALYTIC ACTIVITY: Hydrolyzes a Gly-|-Gly bond at its own C-
CC terminus, commonly in the sequence -Tyr-Xaa-Val-Gly-|-Gly-, in the
CC processing of the potyviral polyprotein.
CC -!- PTM: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.
CC -!- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE
CC POLYPEPTIDE WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC
CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT
CC INDIVIDUAL PROTEINS.
CC -!- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.
CC -!- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.
CC -!- SIMILARITY: BELONGS TO THE POTYVIRUS POLYPEPTIDE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D10930; BAA01726.1; -.
CC PIR; JQ1331; GNVSEFV.
CC MEROPS; C04.010; -.
CC MEROPS; C06.001; -.
CC InterPro; IPR001410; DEAD.
CC InterPro; IPR001450; Helicase_C.
CC InterPro; IPR001720; Peptidase_C4.
CC InterPro; IPR001456; Peptidase_C6.
CC InterPro; IPR001592; Poty_coat.
CC InterPro; IPR002540; Poty_P1.
CC InterPro; IPR007095; RNA_pol_PS.
CC InterPro; IPR001205; RNA_pol_P3.
CC InterPro; IPR007094; RNA_pol_P5vir.
CC Pfam; PF00271; helicase_C.1.
CC Pfam; PF00863; Peptidase_C4.1.
CC Pfam; PF00851; Peptidase_C6.1.
CC Pfam; PF00767; Poty_coat.1.
CC Pfam; PF01577; Poty_P1.1.
CC Pfam; PF00680; RNA_dep_RNA_pol.1.
CC PRINTS; PR00966; NIAPOTYPEASE.
CC SMART; SM00487; DEXD; 1.
CC SMART; SM00490; HELICC; 1.
CC Hydrolase; Transferase; Thiol protease; RNA-directed RNA polymerase;
CC Coat protein; Polyprotein; Covalent protein-RNA linkage; Helicase;
CC ATP-binding.
CC FT CHAIN 1 ? N-TERMINAL PROTEIN.
CC FT CHAIN 2 ? HELPER COMPONENT PROTEINASE.
CC FT CHAIN 857 ? PROTEIN P3.
CC FT CHAIN ? 1266 6 kDa PROTEIN 1.
CC FT CHAIN 1267 1902 CYTOPLASMIC INCLUSION PROTEIN.
CC FT CHAIN 1903 1955 6 kDa PROTEIN 2.
CC FT CHAIN 1956 ? GENOME-LINKED PROTEIN.
CC FT CHAIN ? 2395 ? NUCLEAR INCLUSION PROTEIN A.
CC FT CHAIN 2396 2915 NUCLEAR INCLUSION PROTEIN B.
CC FT CHAIN 2916 3206 COAT PROTEIN.
CC FT CHAIN 1266 1267 CLEAVAGE (BY 49 kDa PROTEASE).
CC FT SITE 1902 1903 CLEAVAGE (BY 49 kDa PROTEASE).
CC FT SITE 1955 1956 CLEAVAGE (BY 49 kDa PROTEASE).
CC FT SITE 2395 2396 CLEAVAGE (BY 49 kDa PROTEASE).
CC FT SITE 2915 2916 CLEAVAGE (BY 49 kDa PROTEASE).
CC FT BINDING 2016 2016 COVALENT LINKAGE OF VIRAL RNA (BY
CC SIMILARITY).
CC FT NP_BIND 1351 1358 ATP (POTENTIAL).
CC SQ SEQUENCE 3206 AA; 364271 MW; 42A3D921BEE9A0CBF CRC64;
CC -----
CC Query Match 57.6%; Score 38; DB 1; Length 3206;
CC Best Local Similarity 50.0%; Pred. No. 1.5e+02;
CC Matches 5; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
CC -----
CC QY 2 PVLHVALNTP 13
CC D6 67 PINHVDSTKPI 78
```

```
RESULT 8
YM21_YEAST STANDARD; PRT; 102 AA.
ID YN21_YEAST
AC F40211;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothesized 12.1 kDa protein in MDS1-RPL13B intergenic region.
DE YMR141C OR YMR375.10C.
GN Saccharomyces cerevisiae (Baker's yeast).
OC Saccharomycetes: Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
CX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=288C / AB972;
RX PubMed=9169872;
RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT XIII."
RL Nature 387:90-93 (1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z47071; CAA87355.1; -.
CC PIR; S50397; S50397.
CC SGD; S0004749; YMR141C.
CC KW Hypothetical protein.
CC SQ SEQUENCE 102 AA; 12108 MW; 8A7F6449CC86F219 CRC64;
CC -----
CC Query Match 56.1%; Score 37; DB 1; Length 102;
CC Best Local Similarity 50.0%; Pred. No. 5.3;
CC Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
CC -----
CC QY 3 VHLHVALNTP 12
CC D6 29 IMHLISNTP 38
CC -----
RESULT 9
EAB_EBV STANDARD; PRT; 191 AA.
ID EAB_EBV
AC P03182;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Early antigen protein R (EA-R) (Nuclear antigen).
DE BHFR1.
GN Epstein-barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
CX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Barrell B.G.;
RA "DNA sequence and expression of the B95-8 Epstein-Barr virus genome."
RL Nature 310:207-211 (1984).
RN [2]
RP SEQUENCE FROM N.A.
```

```

RX MEDLINE=87284169; PubMed=3039177;
RA Pfitzner A.J., Tsai E.C., Strominger J.L., Speck S.H.;
RT "Isolation and characterization of cDNA clones corresponding to
RT transcripts from the BamHI H and F regions of the Epstein-Barr virus
RL genome.";
RL J. Virol. 61:2902-2909 (1987).
RN [3]
RP IDENTIFICATION OF PROTEIN.
RA MEDLINE=87321096; PubMed=2820125;
RA Pearson G.R., Luka J., Petti L., Sample J., Birkenbach M., Braun D.,
RA Kieff E.;
RT "Identification of an Epstein-Barr virus early gene encoding a second
RT component of the restricted early antigen complex.";
RL Virol. 160:151-161 (1987).
CC -!- FUNCTION: INHIBITS APOPTOSIS.
CC -!- MISCELLANEOUS: EA-R IS PART OF THE RESTRICTED EA-COMPLEX.
CC -!- SIMILARITY: Contains 1 Bcl-2 homology 1 (BH1) domain.
CC -!- SIMILARITY: Contains 1 Bcl-2 homology 2 (BH2) domain.
CC -!- SIMILARITY: BELONGS TO THE BCL-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17416; AAA45873.1; -
DR EMBL; V01555; -; NOT ANNOTATED_CDS.
DR EMBL; M17293; AAA45875.1; -
DR EMBL; A22899; CAA01638.1; -
DR PIR; C93065; QQB4.
DR InterPro; IPR000712; Bcl2_BH.
DR InterPro; IPR002475; BCL2_family.
DR PROSITE; PS50062; BCL2_FAMILY; 1.
DR PROSITE; PS01080; BH1; FALSE_NEG.
DR PROSITE; PS01258; BH2; 1.
DR Early protein; Transmembrane; Apoptosis.
KW DOMAIN 99 109 BH1.
FT DOMAIN 142 157
FT TRANSMEM 166 186
FT CARBOHYD 22 22
FT CARBOHYD 118 118 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 191 AA; 21893 MW; 8108BCB94F81DC6B CRC64;

Query Match 56.1%; Score 37; DB 1; Length 191;
Best Local Similarity 66.7%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 PVLHLVALNTP 13
Db 27 PVLELAARETPL 38
|||||
RESULT 10
SKI MOUSE STANDARD; PRT; 348 AA.
AC Q60598;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ski oncogene (C-ski) (Fragment).
GN SKI.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/J;
RX MEDLINE=96127473; PubMed=8573720;
RA Namciu S., Lyons G.E., Micales B.K., Heyman H.-C., Colmenares C.,

RA Stavnezer E.;
RT "Enhanced expression of mouse c-ski accompanies terminal skeletal
RT muscle differentiation in vivo and in vitro.";
RL Dyn. 204:291-300 (1995).
CC -!- FUNCTION: May play a role in terminal differentiation of skeletal
CC muscle cells but not in the determination of cells to the myogenic
CC lineage.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: Is expressed in a uniform pattern in all
CC embryonic cells prior to skeletal muscle cell formation in the
CC myotomes of somites. Expression is first upregulated in skeletal
CC muscle at 12 dpc, this upregulation is evident first in body wall
CC muscle and one day later in limb muscles. At 13.5 dpc a most
CC prominent expression is seen in all skeletal muscles. At this
CC stage expression is seen in all other cells and tissues but at
CC lower levels than in skeletal muscle.
CC -!- SIMILARITY: TO SKI PROTEIN OF SLOAN-KETTERING VIRUS AND TO SNO
CC ONCOGENES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14173; AAA39669.1; -
DR MGD; MG1:98310; Ski.
DR GO; GO:005737; C.cytoplasm; IDA.
DR GO; GO:0005634; C.nucleus; IDA.
DR GO; GO:0017053; C.transcriptional repressor complex; IDA.
DR GO; GO:0005515; F.protein binding activity; IPI.
DR GO; GO:0030326; P.limb morphogenesis; IMP.
DR InterPro; IPR003380; Transform_Ski.
DR Pfam; PF02437; Ski_Sno; 1.
DR Proto-oncogene; Nucleic acid; Repeat.
KW NON_TER 348 348
FT SEQUENCE 348 AA; 37869 MW; 4DD73478145D038C CRC64;

Query Match 56.1%; Score 37; DB 1; Length 348;
Best Local Similarity 63.6%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 PVLHLVALNTP 12
Db 70 PVLHLPAIOFP 80
|||||
RESULT 11
SKI HUMAN STANDARD; PRT; 728 AA.
AC P12755;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Ski oncogene (C-ski).
GN SKI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93345144; PubMed=2762147;
RA Nomura N., Sasamoto S., Ishii S., Matsui M., Ishizaki R.;
RT "Isolation of human cDNA clones of ski and the ski-related gene,
RT sno.";
RL Nucleic Acids Res. 17:5489-5500 (1989).
CC -!- FUNCTION: May play a role in terminal differentiation of skeletal
CC muscle cells but not in the determination of cells to the myogenic
CC lineage.
CC -!- SUBUNIT: INTERACTS WITH SMAD2, 3 AND 4.

```



CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- SIMILARITY: TO SKI PROTEIN OF SLOAN-KETTERING VIRUS AND TO SNO  
CC ONCOGENES.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL: X15218, CA33288.1; .  
CC PIR: S06053; TVRUSK.  
CC PDB: 1MR1; 21-JAN-03.  
CC TRANSFAC: T04643; .  
CC Genew: HGNC:10896; SKI.  
CC MIM: 164780; .  
CC DR GO: GO:0005634; C:nucleus; NAS.  
CC DR GO: GO:0030154; P:cell differentiation; NAS.  
CC DR InterPro: IPR003360; TransForm\_Ski.  
CC DR Pfam: PF02437; Ski\_Sno; 1.  
CC DR Proto-oncogene: Coiled coil; Nuclear protein; Repeat; 3D-structure.  
CC KW COILED COIL (POTENTIAL).  
CC FT DOMAIN 536 710  
CC SQ SEQUENCE 728 AA; 80004 MW; 957804840A28C2DA CRC64;  
Query Match 56.1%; Score 37; DB 1; Length 728;  
Best Local Similarity 63.6%; Pred. No. 45;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 2 PVLHLVALNTP 12  
DB 72 PVLHLPAIOPP 82  
RESULT 12  
VF13-BPH6 STANDARD; PRT; 72 AA.  
AC P1130;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 01-JAN-1990 (Rel. 13, Last annotation update)  
DE P13 protein.  
GN P13.  
OS Bacteriophage phi-6.  
OC Viruses; dsRNA viruses; Cystoviridae; Cystovirus.  
OX NCBI\_TaxID=10879;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.  
RX MEDLINE=88160044; PubMed=3247397;  
RA Gottlieb P., Metzger S., Romantschuk M., Carton J., Strassman J.,  
RA Bamford D.H., Kalkinen N., Mindich L.;  
RT "Nucleotide sequence of the middle dsRNA segment of bacteriophage phi  
RT 6: placement of the genes of membrane-associated proteins."  
RL Virology 163:183-190(1988).  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL: M17462; AAA68486.1; .  
CC PIR: D28648; PNBPF6.  
CC KW Envelope protein.  
SQ SEQUENCE 72 AA; 7649 MW; 61DBAB3B71053B80 CRC64;  
Query Match 54.5%; Score 36; DB 1; Length 72;  
Best Local Similarity 58.3%; Pred. No. 5.5;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTP 12  
DB 14 QPLVLVALNTP 25  
RESULT 13  
UCP4 HUMAN STANDARD; PRT; 323 AA.  
ID UCP4 HUMAN  
AC Q95647;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Mitochondrial uncoupling protein 4 (UCP 4).  
GN UCP4.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=99148824; PubMed=10025957;  
RA Mao W., Yu X.X., Zhong A., Li W., Brush J., Sherwood S.W., Adams S.H.,  
RA Pan G.;  
RT "UCP4, a novel brain-specific mitochondrial protein that reduces  
RT membrane potential in mammalian cells."  
RL FEBS Lett. 443:326-330(1999).  
CC -!- FUNCTION: UCP ARE MITOCHONDRIAL TRANSPORTER PROTEINS THAT CREATE  
CC PROTON LEAKS ACROSS THE INNER MITOCHONDRIAL MEMBRANE, THUS  
CC UNCOUPLING OXIDATIVE PHOSPHORYLATION FROM ATP SYNTHESIS. AS A  
CC RESULT, ENERGY IS DISSIPATED IN THE FORM OF HEAT. MAY PLAY A ROLE  
CC IN THERMOREGULATORY HEAT PRODUCTION AND METABOLISM IN BRAIN.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial  
CC inner membrane.  
CC -!- TISSUE SPECIFICITY: FOUND IN ADULT AND FETAL BRAIN. PRESENT IN  
CC MOST OF THE BRAIN TISSUES, WITH LOW LEVELS IN SPINAL CHORD, CORPUS  
CC CALLOSUM AND SUBSTANTIA NIGRA.  
CC -!- SIMILARITY: Belongs to the mitochondrial carrier family.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC  
CC EMBL: AF110532; AAD16995.1; .  
CC DR GO: GO:0005739; C:mitochondrion; TAS.  
CC DR GO: GO:0015302; F:uncoupling protein activity; TAS.  
CC DR GO: GO:0006091; F:energy pathways; TAS.  
CC DR InterPro: IPR002030; Mit\_uncoupling.  
CC DR InterPro: IPR001993; Mitoch\_carrier.  
CC DR Pfam: PF00153; mito\_carr; 3.  
CC DR PRINTS: PR00784; MTUNCOUPLING.  
CC DR PROSITE: PS00215; MITOCH\_CARRIER; 2.  
CC DR Mitochondrion; Inner membrane; Repeat; Transmembrane; Transport.  
KW TRANSMEM 23 40  
FT TRANSMEM 88 109  
FT TRANSMEM 127 144  
FT TRANSMEM 195 212  
FT TRANSMEM 229 248  
FT TRANSMEM 288 311  
FT DOMAIN 5 8  
SQ SEQUENCE 323 AA; 36064 MW; 4C54A56BB10333ED CRC64;  
Query Match 54.5%; Score 36; DB 1; Length 323;  
Best Local Similarity 63.6%; Pred. No. 28;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 3 VHLVALNTP 13  
DB 214 VKHVLVALNTP 224

```
RA Kavarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
CC -!- FUNCTION: Responsible for the formation of archaea-specific
CC glycerophosphate backbone of phospholipids, G-1-P, from
CC dihydroxyacetonephosphate (DHAP) (by similarity) (-) =
CC -!- CATALYTIC ACTIVITY: Sn-glycerol-1-phosphate + NAD(P)H ->
CC glycerone phosphate + NAD(P)H.
CC -!- PATHWAY: De novo phospholipid biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCEROL-1-PHOSPHATE DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AP000059; BAA79484.1; -.
CC PIR; H72748; H72748.
CC HAMAP; MF 00497; -.
CC InterPro; IPR002658; DHQ synthase.
CC Pfam; PF01741; DHQ synthase.
CC TIGRPFAM; TIGR00238; TIGR00238.1.
KW Phospholipid biosynthesis; Oxidoreductase; NADP; Complete proteome.
SQ SEQUENCE 370 AA; 39352 MW; 0782087EE9FCBF01 CRC64;

Query Match 54.5%; Score 36; DB 1; Length 370;
Best Local Similarity 70.0%; Pred. No. 33;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 4 LHLVALNTPL 13
|:|:|:|:|
Db 277 LHIVARNKPL 286

Search completed: January 26, 2004, 19:09:31
Job time : 27 secs

RA Kavarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
CC -!- FUNCTION: Responsible for the formation of archaea-specific
CC glycerophosphate backbone of phospholipids, G-1-P, from
CC dihydroxyacetonephosphate (DHAP) (by similarity) (-) =
CC -!- CATALYTIC ACTIVITY: Sn-glycerol-1-phosphate + NAD(P)H ->
CC glycerone phosphate + NAD(P)H.
CC -!- PATHWAY: De novo phospholipid biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCEROL-1-PHOSPHATE DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; UI4003; AAA97045.1; -.
CC EMBL; A5000487; AAC77106.1; -.
CC PIR; S56374; S56374.
CC EcoGene; EGI2473; YJek.
CC InterPro; IPR003739; DUF160.
CC Pfam; PF04055; Radical SAM.
CC TIGRPFAM; TIGR00238; TIGR00238.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 342 AA; 38749 MW; C35F066281588CD3 CRC64;

Query Match 54.5%; Score 36; DB 1; Length 342;
Best Local Similarity 75.0%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 5 HLVALNTPL 12
|:|:|:|:|
Db 3 HIVTLNTP 10

RESULT 15
EGSA_AERPE STANDARD; PRT; 370 AA.
AC Q3YER2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Glycerol-1-phosphate dehydrogenase [NAD(P)] (EC 1.1.1.261) (Sn-
DE glycerol-1-phosphate dehydrogenase) (G-1-P dehydrogenase)
DE (Enantiomeric glycerophosphate synthase).
GN EGSA OR APE0519.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KJ;
RX MEDLINE=99310339; PubMed=10382966;

RA Kavarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
CC -!- FUNCTION: Responsible for the formation of archaea-specific
CC glycerophosphate backbone of phospholipids, G-1-P, from
CC dihydroxyacetonephosphate (DHAP) (by similarity) (-) =
CC -!- CATALYTIC ACTIVITY: Sn-glycerol-1-phosphate + NAD(P)H ->
CC glycerone phosphate + NAD(P)H.
CC -!- PATHWAY: De novo phospholipid biosynthesis.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE GLYCEROL-1-PHOSPHATE DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; UI4003; AAA97045.1; -.
CC EMBL; A5000487; AAC77106.1; -.
CC PIR; S56374; S56374.
CC EcoGene; EGI2473; YJek.
CC InterPro; IPR003739; DUF160.
CC Pfam; PF04055; Radical SAM.
CC TIGRPFAM; TIGR00238; TIGR00238.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 342 AA; 38749 MW; C35F066281588CD3 CRC64;
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 26, 2004, 19:04:49 ; Search time 83 Seconds

(without alignments)  
40.418 Million cell updates/sec

Title: US-09-766-412-30

Perfect score: 66  
Sequence: 1 QPVLHLVALNTPL 13

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL 23.\*

- 1: sp\_archea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriaph.\*
- 17: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	66	100.0	184	11 Q9JK63	Q9JK63 mus musculus
2	66	100.0	1140	11 Q61434	Q61434 mus musculus
3	66	100.0	1774	11 Q82001	Q82001 mus musculus
4	62	93.9	187	4 Q8W15	Q8W15 homo sapien
5	62	93.9	261	4 Q8NG19	Q8NG19 homo sapien
6	62	93.9	816	4 Q8N4S4	Q8N4S4 homo sapien
7	62	93.9	1344	13 Q93419	Q93419 gallus gall
8	61	92.4	226	11 Q9QZD2	Q9QZD2 rattus norv
9	59	89.4	1367	11 Q9EOD9	Q9EOD9 mus musculus
10	59	89.4	1367	11 Q35206	Q35206 mus musculus
11	54	81.8	171	11 Q9WUW5	Q9WUW5 rattus norv
12	52	78.8	1315	13 Q8QHL9	Q8QHL9 xenopus lae
13	50	75.8	1307	13 Q8JFF7	Q8JFF7 xenopus lae
14	45	68.2	217	8 Q9B6Q6	Q9B6Q6 eriosoma la
15	45	68.2	361	13 Q8AMC6	Q8AMC6 brachydanio
16	45	68.2	1388	4 Q9Y4W4	Q9Y4W4 homo sapien

#### RESULT 1

Q9JK63 ID Q9JK63 PRELIMINARY; PRT; 184 AA.  
AC Q9JK63; 2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 15, Last annotation update)  
DE Endostatin (Fragment).  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_taxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Chinese Kunming;  
RA Jia S., Zhu F., Xing G., Yu Y., Duan C., Xiu R.-J., He F.;  
RT "Anticancer treatment of targeted fusion protein delivery to tumor neovasculation."  
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF257775; AAF69009.1; -.  
DR HSSP; P39061; 1KOE.  
FT NON\_TER 1 184  
SQ SEQUENCE 184 AA; 20376 MW; AC06F9D8D103412A CRC64;  
Query Match 100.0%; Score 66; DB 11; Length 184;  
Best Local Similarity 100.0%; Pred. No. 0.00025;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

#### RESULT 2

Q61434 ID Q61434 PRELIMINARY; PRT; 1140 AA.  
AC Q61434;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)

17 43 65.2 217 8 Q9B6P8  
18 43 65.2 218 8 Q9B6G9  
19 41 62.1 568 11 Q9Z1N7  
20 40 60.6 176 16 O05584  
21 40 60.6 217 8 Q9B7T1  
22 40 60.6 217 8 Q9B6I2  
23 40 60.6 217 8 Q9B7O6  
24 40 60.6 217 8 Q9B6Q8  
25 40 60.6 217 8 Q9B6H4  
26 40 60.6 217 8 Q9B6H5  
27 40 60.6 217 8 Q9B6K0  
28 40 60.6 217 8 Q9B6G3  
29 40 60.6 217 8 Q9B6Q1  
30 40 60.6 217 8 Q9B6I5  
31 40 60.6 217 8 Q9B6H9  
32 40 60.6 217 8 Q9B6I0  
33 40 60.6 217 8 Q9B6Q0  
34 40 60.6 217 8 Q9B6H8  
35 40 60.6 217 8 Q9B6R3  
36 40 60.6 217 8 Q9B6J3  
37 40 60.6 217 8 Q9B6G4  
38 40 60.6 342 16 Q8ZKB8  
39 40 60.6 342 16 Q8Z1A2  
40 40 60.6 1538 10 Q94H26  
41 39 59.1 268 17 Q8TW56  
42 39 59.1 333 2 Q46143  
43 39 59.1 435 16 Q8Z530  
44 39 59.1 434 4 Q96HR8  
45 39 59.1 533 10 Q9LPS5

#### ALIGNMENTS

Q9B6P8 geioica utri  
Q9B6G9 tetraneura  
Q9Z1N7 mus musculu  
O05584 mycobacteri  
Q9B7T1 apioneura l  
Q9B6I2 panaphis ju  
Q9B7O6 acyrthosiph  
Q9B6Q8 drepanosiph  
Q9B6H4 schizolachn  
Q9B6H5 schizaphis  
Q9B6K0 lachnus rob  
Q9B6G3 thelaxus su  
Q9B6Q1 eulachnus s  
Q9B6I5 pemphigus b  
Q9B6H9 phylloxera  
Q9B6I0 pemphigus s  
Q9B6Q0 forda formi  
Q9B6H8 phylloxera  
Q9B6R3 chaetophoru  
Q9B6J3 myzus persi  
Q9B6G4 tuberolachn  
Q8ZKB8 salmonella  
Q8Z1A2 salmonella  
Q94H26 cryza sativ  
Q8TW56 methanosarc  
Q46143 clostridium  
Q8Z530 salmonella  
Q96HR8 homo sapien  
Q9LPS5 arabidopsis

```

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Collagen (Fragment).
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Abe N., Muragaki Y., Yoshioka H., Inoue H., Ninomiya Y.;
RT "Identification of a novel collagen chain represented by extensive
RT interruptions in the triple-helical region.";
RL Cell. Mol. Biol. Res. 196:576-582(1993).
DR EMBL; D15461; BAA04483.1; -.
DR HSSP; P39061; 1KOE.
DR MGD; MGI:88451; Coll18a1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 8.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 1.
KW Collagen.
FT NON TER
SQ SEQUENCE 1140 AA; 115156 MW; 8B0C7E6862B3BDFE CRC64;

Query Match 100.0%; Score 66; DB 11; Length 1140;
Best Local Similarity 100.0%; Pred. No. 0.0016;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
DB 963 QPVLHLVALNTPL 975

RESULT 3
Q62001 PRELIMINARY; PRT; 1774 AA.
AC Q62001; Q60672;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Procollagen, type XVIII, alpha 1 precursor (XVIII) collagen
DE (Procollagen, type XVIII, alpha 1) (Alpha-1 type XVIII collagen).
GN COL18A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rehn M., Hintikka E., Pihlajaniemi T.;
RT "Primary structure of the alpha 1 chain of mouse type XVIII collagen,
RT partial structure of the corresponding gene, and comparison of the
RT alpha 1 (XVIII) chain with its homologue, the alpha 1(XV) collagen
RT chain.";
RL J. Biol. Chem. 269:13929-13935(1994).
RN [2]
RP SEQUENCE OF 1-562 FROM N.A.
RA Rehn M., Pihlajaniemi T.;
RT "Alpha 1 (XVIII), a collagen chain with frequent interruptions in the
RT collagenous sequence, a distinct tissue distribution, and homology
RT with type XV collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).
RN [3]
RP SEQUENCE OF 1-562 FROM N.A.
RA Rehn M., Pihlajaniemi T.;
RT "Identification of three N-terminal ends of type XVIII collagen chains
RT and tissue-specific differences in the expression of the corresponding

```

```

RT transcripts. The longest form contains a novel motif homologous to rat
RT and Drosophila frizzled proteins";
RL J. Biol. Chem. 270:4705-4711(1995).
DR EMBL; U03715; AAC52903.1; -.
DR EMBL; U03716; AAC52903.1; JOINED.
DR EMBL; U03718; AAC52903.1; JOINED.
DR EMBL; U34607; AAC52903.1; JOINED.
DR EMBL; U34608; AAC52903.1; JOINED.
DR EMBL; U34609; AAC52903.1; JOINED.
DR EMBL; U34610; AAC52903.1; JOINED.
DR EMBL; U34611; AAC52903.1; JOINED.
DR EMBL; U34612; AAC52903.1; JOINED.
DR EMBL; U34613; AAC52903.1; JOINED.
DR EMBL; U11637; AAC52179.1; -.
DR HSSP; P39061; 1KOE.
DR MGD; MGI:88451; Coll18a1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 8.
DR Pfam; PF01392; Fz; 1.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 1.
DR SMART; SM00063; Fz; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
DR PROSITE; PS00038; FZ; 1.
KW Collagen; Signal.
SQ SEQUENCE 1774 AA; 182229 MW; CF4D9BC9E88EF232 CRC64;

Query Match 100.0%; Score 66; DB 11; Length 1774;
Best Local Similarity 100.0%; Pred. No. 0.0025;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
DB 1597 QPVLHLVALNTPL 1609

RESULT 4
Q8WXI5 PRELIMINARY; PRT; 187 AA.
AC Q8WXI5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Collagen XVII (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21409408; PubMed=11517600;
RA Feng Y., Cui L.B., Liu C.X., Ma Q.J.;
RT "Inhibition effect in vitro of purified endostatin expressed in Pichia
RT pastoris";
RL Sheng Wu Gong Cheng Xue Bao 17:278-282(2001).
DR EMBL; AF416592; AAL37720.1; -.
FT NON TER
SQ SEQUENCE 187 AA; 20448 MW; 72B1047D85938CD3 CRC64;

Query Match 93.9%; Score 62; DB 4; Length 187;
Best Local Similarity 92.3%; Pred. No. 0.0014;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
DB 11 QPVLHLVALNSPL 23

RESULT 5

```

```

Q8NG19
ID Q8NG19 PRELIMINARY; PRT; 261 AA.
AC Q8NG19;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Multi-functional protein MFP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Dou D.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF282883; AAMS2249.1; -.
SQ SEQUENCE 261 AA; 26745 MW; CA60C920AF3E90E5 CRC64;

Query Match 93.9%; Score 62; DB 4; Length 261;
Best Local Similarity 92.3%; Pred. No. 0.002;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
Db 85 QPVLHLVALNSPL 97

RESULT 6
Q8N4S4
ID Q8N4S4 PRELIMINARY; PRT; 816 AA.
AC Q8N4S4;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Similar to collagen, type XVIII, alpha 1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC033715; AAH33715.1; -.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 6.
DR ProDom; PD000007; Collagen; 1.
KW Collagen.
FT NON_TER.
SQ SEQUENCE 816 AA; 82553 MW; 5D539B2946694F86 CRC64;

Query Match 93.9%; Score 62; DB 4; Length 816;
Best Local Similarity 92.3%; Pred. No. 0.0063;
Matches 12; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
Db 640 QPVLHLVALNSPL 652

RESULT 7
O93419
ID O93419 PRELIMINARY; PRT; 1344 AA.
AC O93419;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-OCT-2001 (T-EMBLrel. 18, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Collagen XVIII precursor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.

```

---

```

OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98411346; PubMed=9738008;
RA Halfter W., Dong S., Schurer B., Cole G.J.;
RT "Collagen XVIII is a basement membrane heparan sulfate proteoglycan."
RL J. Biol. Chem. 273:25404-25412(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Halfter W., Dong S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083440; AAC33294.2; -.
DR HSSP; P39061; 1KOE.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 8.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 2.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen; Signal.
FT SIGNAL.
SQ SEQUENCE 1344 AA; 137402 MW; 7AA366B4FE940CCD CRC64;

Query Match 93.9%; Score 62; DB 13; Length 1344;
Best Local Similarity 92.3%; Pred. No. 0.01;
Matches 12; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QPVLHLVALNTPL 13
Db 1167 QPALHLVALNTPL 1179

RESULT 8
Q9QZD2
ID Q9QZD2 PRELIMINARY; PRT; 226 AA.
AC Q9QZD2;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (T-EMBLrel. 22, Last annotation update)
DE Collagen XVIII (Fragment).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=20227226; PubMed=10766159;
RA Perletti G., Concarl P., Giardini R., Marras E., Piccinini F.,
RA Folkman J., Chen L.;
RT "Antitumor activity of endostatin against carcinogen-induced rat
RT primary mammary tumors."
RL Cancer Res. 60:1793-1796(2000).
DR EMBL; AF189709; AAF00975.1; -.
DR HSSP; P39061; 1KOE.
FT NON_TER.
SQ SEQUENCE 226 AA; 25350 MW; 38B83C0486C0E949 CRC64;

Query Match 92.4%; Score 61; DB 11; Length 226;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PVLHLVALNTPL 13
Db 50 PVLHLVALNTPL 61

RESULT 9
Q9EQD9
ID Q9EQD9 PRELIMINARY; PRT; 1367 AA.
AC Q9EQD9;

```

```

DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type XV collagen.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97480713; PubMed=9339358;
RA Haeg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains."
RL Genomics 45:31-41(1997).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=20522048; PubMed=11069303;
RA Eklund L., Mouna A., Luetard J., Pihlajaniemi T.;
RT "Structure of the mouse type XV collagen gene, Coll5a1, comparison
RT with the human COL15A1 gene and functional analysis of the promoters
RT of both genes."
RL Matrix Biol. 19:489-500(2000).
DR ENBL; AP261131; AAG27545.1; JOINED.
DR ENBL; AP261109; AAG27545.1; JOINED.
DR ENBL; AP261110; AAG27545.1; JOINED.
DR ENBL; AP261111; AAG27545.1; JOINED.
DR ENBL; AP261112; AAG27545.1; JOINED.
DR ENBL; AP261113; AAG27545.1; JOINED.
DR ENBL; AP261114; AAG27545.1; JOINED.
DR ENBL; AP261115; AAG27545.1; JOINED.
DR ENBL; AP261116; AAG27545.1; JOINED.
DR ENBL; AP261117; AAG27545.1; JOINED.
DR ENBL; AP261118; AAG27545.1; JOINED.
DR ENBL; AP261119; AAG27545.1; JOINED.
DR ENBL; AP261120; AAG27545.1; JOINED.
DR ENBL; AP261121; AAG27545.1; JOINED.
DR ENBL; AP261122; AAG27545.1; JOINED.
DR ENBL; AP261123; AAG27545.1; JOINED.
DR ENBL; AP261124; AAG27545.1; JOINED.
DR ENBL; AP261125; AAG27545.1; JOINED.
DR ENBL; AP261126; AAG27545.1; JOINED.
DR ENBL; AP261127; AAG27545.1; JOINED.
DR ENBL; AP261128; AAG27545.1; JOINED.
DR ENBL; AP261129; AAG27545.1; JOINED.
DR ENBL; AP261130; AAG27545.1; JOINED.
DR HSSP; P39061; 1KOE.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF01391; Collagen; 5.
DR ProDom; PD000007; Collagen; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1367 AA; 140454 MW; 40F259A3CD47C982 CRC64;
Query Match 89.4%; Score 59; DB 11; Length 1367;
Best Local Similarity 84.6%; Pred. No. 0.039;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 QPVHLVALNTPL 13
:|||||
Db 1194 RPYHLVALNTPV 1206

```

```

RESULT 10
Q35206
ID C35206 PRELIMINARY; PRT; 1367 AA.
AC C35206;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type XV collagen.
GN COL15A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97480713; PubMed=9339358;
RA Haeg P.M., Horelli-Kuitunen N., Eklund L., Palotie A.,
RA Pihlajaniemi T.;
RT "Cloning of mouse type XV collagen sequences and mapping of the
RT (XV) collagen sequences indicates divergence in the number of small
RT collagenous domains."
RL Genomics 45:31-41(1997).
DR ENBL; AF011450; AAC53387.1; -.
DR HSSP; P39061; 1KOE.
DR MGD; MGI:88449; Coll5a1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001791; Laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 5.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1367 AA; 140525 MW; A483A1254AF3ABEC CRC64;
Query Match 89.4%; Score 59; DB 11; Length 1367;
Best Local Similarity 84.6%; Pred. No. 0.039;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 QPVHLVALNTPL 13
:|||||
Db 1194 RPYHLVALNTPV 1206

```

```

RESULT 11
Q9WUW5
ID Q9WUW5 PRELIMINARY; PRT; 171 AA.
AC Q9WUW5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Collagen type XVIII, alpha (I) chain (Fragment).
GN COL18A1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Jia J.D., Bauer M., Eberspaecher U., Donner P., Schuppan D.;
RT "Partial 3'-terminal cDNA sequence of rat collagen XVIII/endostatin."
RT Submitted (FEB-1999) to the ENBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Jia J.D., Bauer M., Sedlaczek N., Ruchl M., Riecken E.O., Schuppan D.;
RT "Temporopartial expression of collagen XVIII/endostatin in acute and
RT chronic liver injuries."
RT Submitted (FEB-1999) to the ENBL/GenBank/DBJ databases.
DR ENBL; A023687; CAB44263.1; -.

```

```

DR HSP: P39061; 1KOE.
DR NON TER 1
FT NON TER 171 171
SQ SEQUENCE 171 AA; 18933 MW; 81BE2EB3FC2C8E72 CRC64;
Query Match 81.8%; Score 54; DB 11; Length 171;
Best Local Similarity 100.0%; Pred. No. 0.04; 0; Indels 0; Gaps 0;
Matches 11; Conservative 0; Mismatches 0;
QY 3 VLHLVALNTPL 13
DB 1 VLHLVALNTPL 11
RESULT 12
Q8QHL9 PRELIMINARY; PRT; 1315 AA.
AC Q8QHL9
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type XVII collagen alpha chain.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Ishino T., Sekimizu K., Natori S., Kubo T.;
RT "Identification and characterization of genes expressed selectively in
RT the regenerating tail of Xenopus laevis tadpole.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047066; BAB84674.1; -.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF03391; Collagen; 7.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Collagen; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1315 AA; 134946 MW; 0C56C235DE058365 CRC64;
Query Match 78.8%; Score 52; DB 13; Length 1315;
Best Local Similarity 83.3%; Pred. No. 0.75; 2; Indels 0; Gaps 0;
Matches 10; Conservative 0; Mismatches 0;
QY 2 PVLHLVALNTPL 13
DB 1139 PALHLVALNAPL 1150
RESULT 13
Q8JFF7 PRELIMINARY; PRT; 1307 AA.
AC Q8JFF7
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Type XVII collagen short variant.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22166979; PubMed=12175494;
RA Elamra H., Peterson J., Pihlajaniemi T., Destree O.;
RT "Cloning of three variants of type XVII collagen and their expression
RT patterns during Xenopus laevis development.";
RL Mech. Dev. 114:109-113(2002).
DR EMBL; AY052763; AAL14257.1; -.

```

```

DR InterPro; IPR000087; Collagen.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01391; Collagen; 6.
DR Pfam; PF02210; TSPN; 1.
DR SMART; SM00210; TSPN; 1.
KW Collagen.
SQ SEQUENCE 1307 AA; 134153 MW; D53EDBFE3DE34976 CRC64;
Query Match 75.8%; Score 50; DB 13; Length 1307;
Best Local Similarity 75.0%; Pred. No. 1.8; 1; Indels 0; Gaps 0;
Matches 9; Conservative 1; Mismatches 2;
QY 2 PVLHLVALNTPL 13
DB 1131 PALHLVALNAPI 1142
RESULT 14
Q9B6Q6 PRELIMINARY; PRT; 217 AA.
ID Q9B6Q6
AC Q9B6Q6
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ATP synthase A chain subunit 6 (EC 3.6.3.14).
OS Eriosoma lanuginosum.
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Paraneoptera; Hemiptera; Sternorrhyncha; Aphidiformes;
OC Aphidoidea; Pemphigidae; Eriosomatinae; Eriosoma.
OX NCBI_TaxID=135953;
RN [1]
RP SEQUENCE FROM N.A.
RA Martinez-Torres D., Buades C., Latorre A., Moya A.;
RT "Molecular systematics of aphids and their primary endosymbionts.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
CC (BY SIMILARITY).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR EMBL; AJ298665; CAC27833.1; -.
DR InterPro; IPR000568; ATPsynth_Asub.
DR Pfam; PF00119; ATP-synt_A; 1.
DR PRINTS; PR00123; ATPASEA.
DR TIGRfams; TIGR01131; ATP_synth_6_or_A; 1.
DR CF(0); Hydrogen ion transport; Ion transport; Transmembrane;
KW Transport; Mitochondrion.
SQ SEQUENCE 217 AA; 25552 MW; DFFB88F5D5AD49C8 CRC64;
Query Match 58.2%; Score 45; DB 8; Length 217;
Best Local Similarity 63.6%; Pred. No. 2.5; 1; Indels 0; Gaps 0;
Matches 7; Conservative 3; Mismatches 1;
QY 3 VLHLVALNTPL 13
DB 127 IHLPLNTPL 137
RESULT 15
Q9AWC6 PRELIMINARY; PRT; 361 AA.
ID Q9AWC6
AC Q9AWC6
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Collagen XVIII (Fragment).
GN COL18A1.

```

```

OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
CX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Hattak Z., Morvan-Dubois G., Thisse B., Garrone R., Le Guellec D.;
RT "Sequence and embryonic expression of collagen XVIII NC11 domain
RT (endostatin) in the zebrafish.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ494837; CAD38825.1; -
FT NON TER
FT 1
SQ SEQUENCE 361 AA; 40222 MW; 3C5A0F8479D26735 CRC64;

Query Match 68.2%; Score 45; DB 13; Length 361;
Best Local Similarity 72.7%; Pred. No. 4.1;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 2 PVLHLVALNTP 12
Db 185 PGLHLIALNSP 195
| | | | | | | |
| | | | | | | |
| | | | | | | |

Search completed: January 26, 2004, 15:08:03
Job time : 88 secs

```